PBR Diffuse Lighting for GGX+Smith Microsurfaces

Earl Hammon, Jr.
Lead Software Engineer
Respawn Entertainment, LLC

Who am I? Well, my name is Earl…

GGX diffuse research done during Titanfall 2 development

Key Takeaways

- New diffuse equations derived from same assumptions as GGX+Smith specular
- New cheap and good shadowing/masking function G for GGX+Smith specular
- New trig identities for shader optimization

Some fun discoveries on the way

- Why’s that “4” in PBR specular?
- What is “s” in Oren Nayar diffuse?
- Smith shadowing/masking assumptions
- A physical interpretation of Lambert
- Help interpreting Disney’s BRDF slices

Quick aside: Original motivation

- Titanfall 2 used Oren-Nayar diffuse
- Question: How to get Oren-Nayar's roughness s from GGX's roughness α?
- Discovery: Oren-Nayar came from very different assumptions!

Quick aside: Original motivation

	Oren-Nayar	**GGX+Smith**
Shadowing/Masking	V-cavities	Smith
Normal Distribution	Spherical Gaussian	GGX
Roughness parameter	$s \in [0, \infty]$	$\alpha \in [0,1]$
Perfectly flat	$s = 0$	$\alpha = 0$
Standard deviation of slopes of normals	s	$0 \quad \alpha = 0$ $\alpha^2\infty \quad \alpha \neq 0$

Quick aside: Original motivation

- Oren-Nayar and Smith+GGX don’t match!
  - Can’t even match standard deviations
  - Hmm… GGX standard deviation is $\alpha^2\infty$
    - Maybe “best” to mipmap/filter α^2?
    - Sum of two GGX distributions is not GGX, so can’t mipmap/filter “properly”

Road map for today’s talk

- General microfacet-based BRDFs
- Simulating diffuse for GGX+Smith microfacet model
- Comparing to other diffuse BRDFs

Road map for today’s talk

- General microfacet-based BRDFs
- Simulating diffuse for GGX+Smith microfacet model
- Comparing to other diffuse BRDFs

Microfacet BRDF sub-topic map

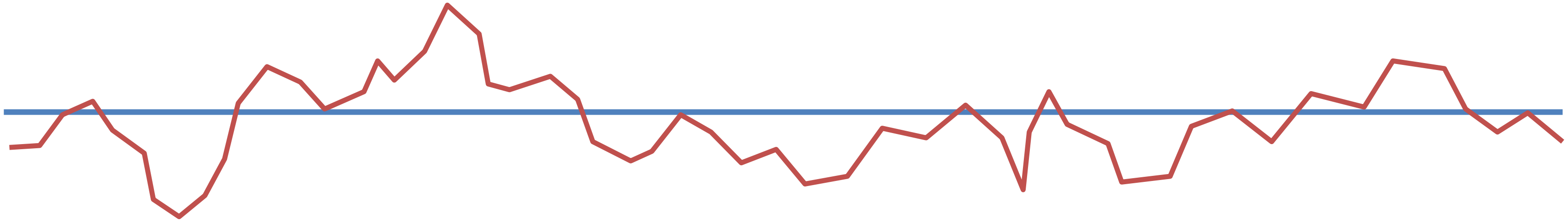
- General form
- How we get PBR specular from that
- Extend to diffuse BRDF

Microfacet BRDF sub-topic map

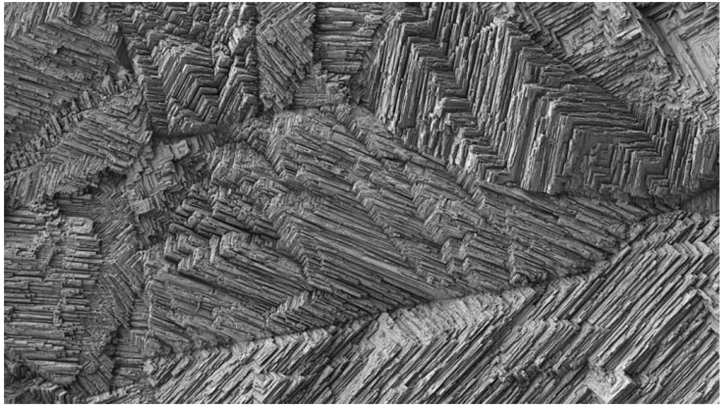
- General form
- How we get PBR specular from that
- Extend to diffuse BRDF

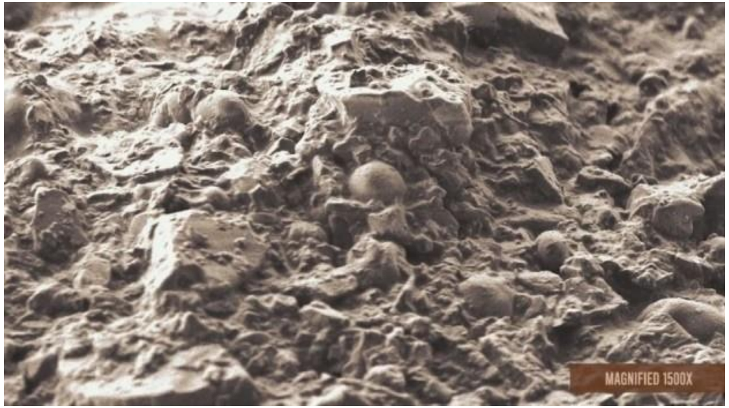
Microfacet models

- Complex macrosurface BRDF averages many microfacets that use a simple BRDF
- Basically just subpixel shader antialiasing



Real world examples

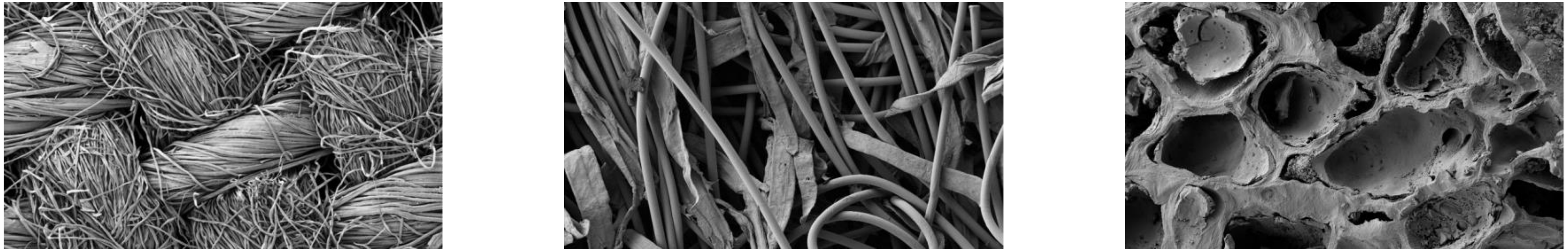




Images: http://funjungle.net/the-world-is-different-under-the-microscope/
http://www.geek.com/news/chocolate-under-an-electron-microscope-looks-like-an-alien-planet-1648301/

“General” Microfacet-based BRDF

- Not fully general; assumes heightfield
  - No weaves, arches, or caves



Images: http://funjungle.net/the-world-is-different-under-the-microscope/

General Microfacet-based BRDF

- $\int_{\Omega} \rho_m(L, V, m) D(m) G_2(L, V, m) \frac{\langle m \cdot L \rangle}{|N \cdot L|} \frac{\langle m \cdot V \rangle}{|N \cdot V|} dm$

General Microfacet-based BRDF

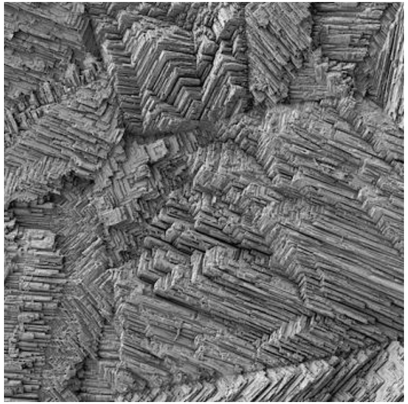
- $\int_\Omega \rho_m(L,V,m)D(m)G_2(L,V,m)\frac{\langle m\cdot L\rangle}{|N\cdot L|}\frac{\langle m\cdot V\rangle}{|N\cdot V|}dm$
    - Integral over all microfacet normals

General Microfacet-based BRDF

- $\int_{\Omega} \rho_m(L,V,m) D(m) G_2(L,V,m) \frac{\langle m\cdot L\rangle}{|N\cdot L|} \frac{\langle m\cdot V\rangle}{|N\cdot V|} dm$
  - How an individual facet responds to light
    - I.e., microfacet BRDF centered on m
    - Usually ideal mirror or ideal diffuse

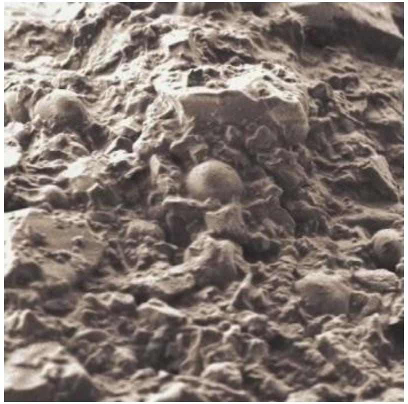
General Microfacet-based BRDF

- $\int_\Omega \rho_m(L,V,m)D(m)G_2(L,V,m)\frac{\langle m\cdot L\rangle}{|N\cdot L|}\frac{\langle m\cdot V\rangle}{|N\cdot V|}dm$
  - Probability density of normal m
  - Which facet normals exist, but not their arrangement (shape)



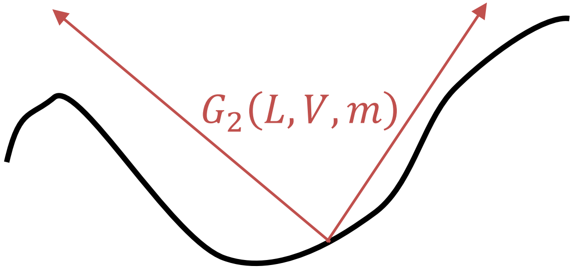
General Microfacet-based BRDF

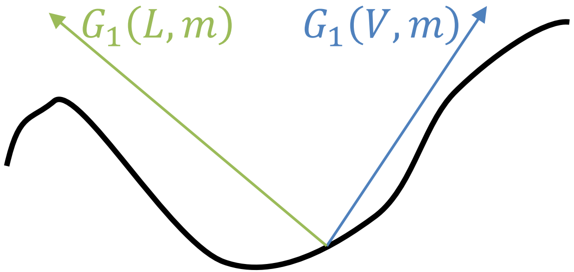
- $\int_\Omega \rho_m(L,V,m)D(m)G_2(L,V,m)\frac{\langle m\cdot L\rangle}{|N\cdot L|}\frac{\langle m\cdot V\rangle}{|N\cdot V|}dm$
  - Occlusion due to actual microfacet arrangement (actual shape)
    - A.k.a. shadowing/masking function
  - Probability microfacet m sees both light L and viewer V



Microfacet BRDF: G_2 vs G_1

- $G_2(L, V, m)$ is % visible in 2 directions
- $G_1(V, m)$ is % visible in just 1 direction
  - In practice, G_2 is derived from G_1





Microfacet BRDF: $D(m)$ properties

- $\int_{\Omega} \rho_m(L,V,m) D(m) G_2(L,V,m) \frac{\langle m \cdot L \rangle}{|N \cdot L|} \frac{\langle m \cdot V \rangle}{|N \cdot V|} dm$
  - Probability density of normal m
    - How quickly cumulative probability changes near m
      - Will change more quickly in more probable regions
    - In range $[0, \infty]$, not $[0,1]$!
      - $D(m) = \infty$ for any m whose probability $\neq 0$!

Microfacet BRDF: $D(m)$ properties

- $\int_\Omega D(m)dm =?$
  - Total surface area of all microfacets
  - Always > 1 if any roughness at all!
- $\int_\Omega D(m) \cos\theta_m \, dm = 1$
  - To normalize total area, project microfacets onto macrosurface using $\cos\theta_m = m \cdot N$

General Microfacet-based BRDF

- $\int_\Omega \rho_m(L,V,m) D(m) G_2(L,V,m) \frac{\langle m \cdot L \rangle}{|N \cdot L|} \frac{\langle m \cdot V \rangle}{|N \cdot V|} dm$
  - Probability density of having microfacet normal m that is both lit and seen
    - I.e., probability density of using BRDF $\rho_m(L,V,m)$.

General Microfacet-based BRDF

- $\int_\Omega \rho_m(L,V,m) D(m) G_2(L,V,m) \frac{\langle m \cdot L \rangle}{|N \cdot L|} \frac{\langle m \cdot V \rangle}{|N \cdot V|} \, dm$
  - $\frac{\langle m \cdot L \rangle}{|N \cdot L|}$ - How big facet m appears to the light
  - $\frac{\langle m \cdot V \rangle}{|N \cdot V|}$ - How big facet m appears to the viewer
    - I.e., normalize contribution from light and to viewer

General Microfacet-based BRDF

- $\int_{\Omega} \rho_m(L,V,m) D(m) G_2(L,V,m) \frac{\langle m \cdot L \rangle}{|N \cdot L|} \frac{\langle m \cdot V \rangle}{|N \cdot V|} dm$
  - Probability density of light from L reaching V in a single bounce off microfacet normal m
  - Requirement: $\int_{\Omega} D(m) G_2(L,V,m) \frac{\langle m \cdot L \rangle}{|N \cdot L|} \frac{\langle m \cdot V \rangle}{|N \cdot V|} dm \leq 1$
    - Only $= 1$ for flat $D(m)$ - **too dark if any roughness!**

General Microfacet-based BRDF

- Related requirement:

$$\int_{\Omega} D(m) G_1(V, m) \langle m \cdot V \rangle dm = |N \cdot V|$$

  - In any direction V, total visible microfacet area equals macrosurface area

Microfacet BRDF sub-topic map

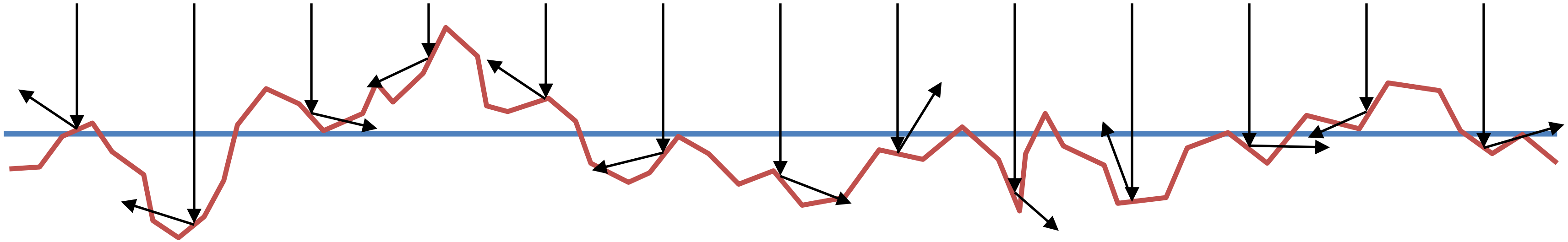
- General form
- How we get PBR specular from that

$$\frac{F(L,H)D(H)G_2(L,V,H)}{4|N \cdot L||N \cdot V|}$$

  - Also: Where does that 4 come from, and why isn’t it π?
- Extend to diffuse BRDF

PBR Specular Microfacet BRDF

- Microfacet BRDF is a perfect mirror
    - I.e., light reflects if and only if $m = H$
        - Mathematically, BRDF is a scaled dirac delta $\delta_m(H, m)$



PBR Specular Microfacet BRDF

- Pure mirror BRDF: $k\delta_m(H,m)$
    - $\delta_m(H,m)$ is the dirac delta using measure m
    - k is some normalization factor we must find
- Normalized BRDF: $\int_\Omega \rho(L,V,N)\cos\theta_V\, dV = 1$
    - For any light and normal, all energy reflects to exactly one viewer

Specular BRDF normalization

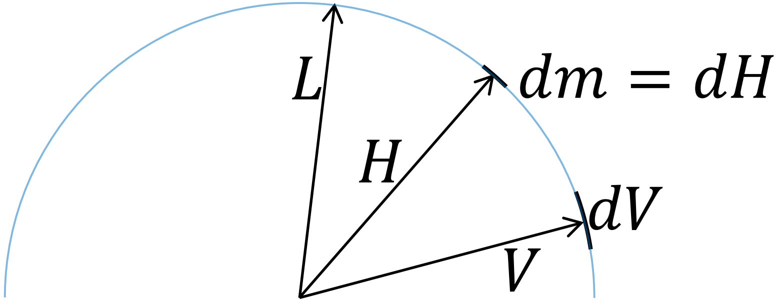
- General case: $\int_{\Omega} \rho(L, V, N) \cos\theta_V \, dV = 1$
- Our case: $\int_{\Omega} k\delta_m(H, m) \cos\theta_V \frac{dV}{dm} dm = 1$
  - Must integrate over dm to evaluate δ_m, so find $\frac{dV}{dm}$ to change integration domain

Specular BRDF normalization

- $\frac{dV}{dm}$ is how fast V changes relative to m
- This will introduce PBR specular's 4!
  - Next few slides show how

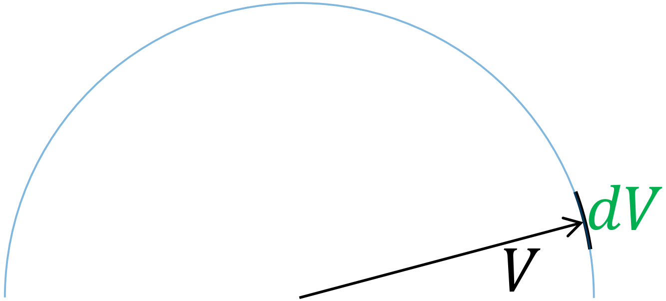
Specular BRDF normalization

- We’re going to find dm from dV to get $\frac{dV}{dm}$
- First, δ_m picks $m = H$, so $dm = dH$
  - All vectors sketched on unit hemisphere



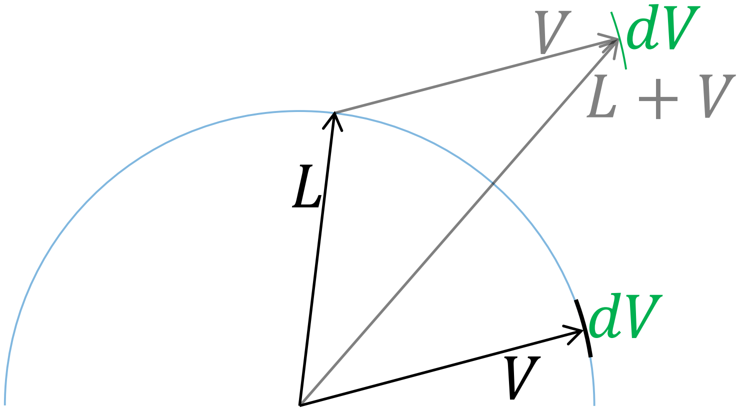
Specular BRDF normalization

- Move solid angle dV...



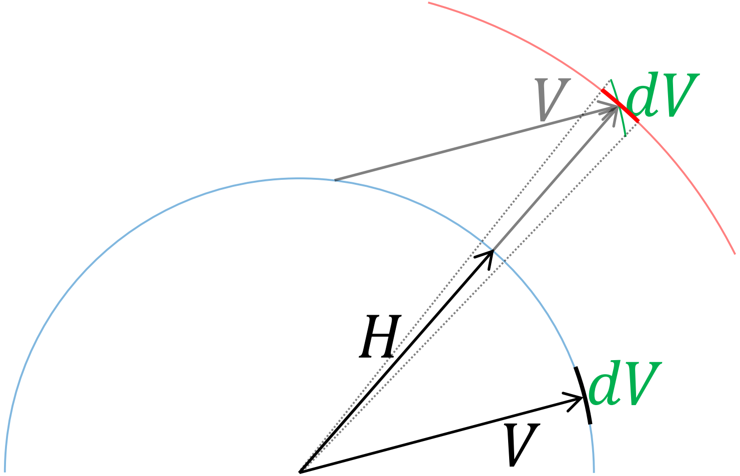
Specular BRDF normalization

- Move solid angle dV to $L + V$...



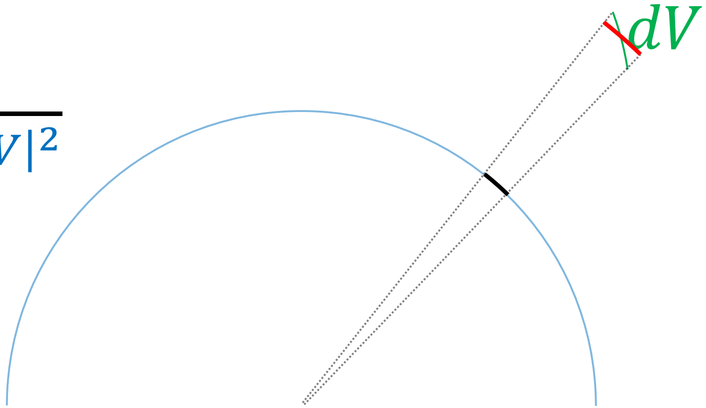
Specular BRDF normalization

- Move solid angle dV to $L + V$, scale by…
  - $L + V$ sphere: $H \cdot V$



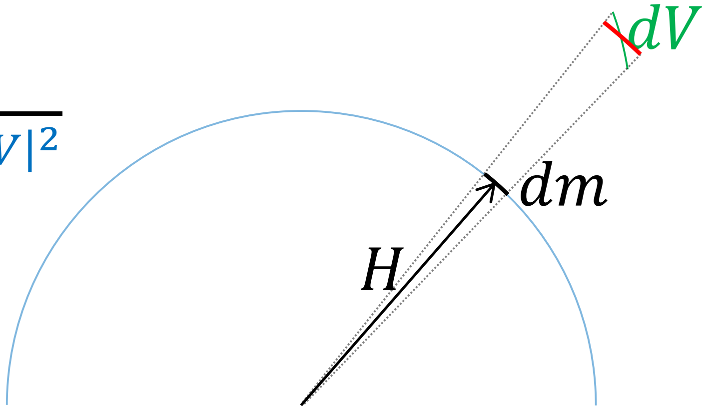
Specular BRDF normalization

- Move solid angle dV to $L+V$, scale by…
  - $L+V$ sphere: $H \cdot V$
  - Unit sphere: $\frac{4\pi\, 1^2}{4\pi |L+V|^2} = \frac{1}{|L+V|^2}$



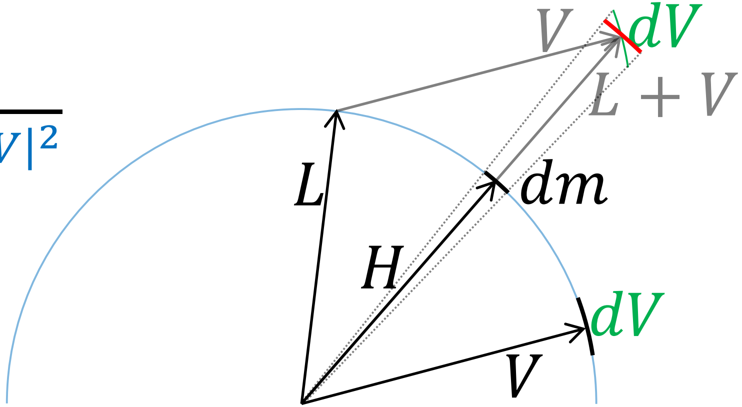
Specular BRDF normalization

- Move solid angle dV to $L+V$, scale by…
    - $L+V$ sphere: $H \cdot V$
    - Unit sphere: $\frac{4\pi\, 1^2}{4\pi |L+V|^2} = \frac{1}{|L+V|^2}$
- $dm = \frac{H \cdot V}{|L+V|^2} dV$



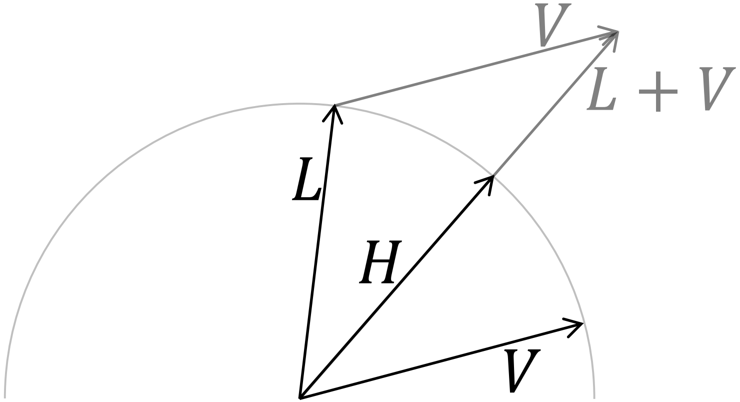
Specular BRDF normalization

- Move solid angle dV to $L + V$, scale by…
  - $L + V$ sphere: $H \cdot V$
  - Unit sphere: $\frac{4\pi\, 1^2}{4\pi |L+V|^2} = \frac{1}{|L+V|^2}$
- $dm = \frac{H \cdot V}{|L+V|^2} dV$



Specular BRDF normalization

- $|L+V| = H \cdot (L+V) = H \cdot L + H \cdot V = 2H \cdot V$
  - This 2 squared is specular BRDF's 4!
- $dm = \frac{H \cdot V}{|L+V|^2} dV = \frac{H \cdot V}{4(H \cdot V)^2} dV$
- $\frac{dV}{dm} = 4H \cdot V$



Specular BRDF normalization

- $\int_\Omega k\delta_m(H,m)\cos\theta_V \frac{dV}{dm}\,dm = 1$
  - $\int_\Omega k\delta_m(H,m)(m\cdot V)(4H\cdot V)dm = 1$
  - $k = \frac{1}{4(H\cdot L)(H\cdot V)}$ since $m = H$ and $H\cdot V = H\cdot L$
- So, pure mirror BRDF: $\frac{\delta_m(H,m)}{4(H\cdot L)(H\cdot V)}$

Specular Microfacet BRDF

- Only the Fresnel reflection fraction $F(L, m)$ of incoming light does specular reflection
- So, final specular microfacet BRDF:

$$\rho_m(L, V, m) = F(L, m)\frac{\delta_H(m)}{4|H \cdot L||H \cdot V|}$$

Specular Microfacet BRDF

- $\int_\Omega \rho_m(L,V,m) D(m) G_2(L,V,m) \frac{\langle m \cdot L \rangle}{|\mathrm{N} \cdot \mathrm{L}|} \frac{\langle m \cdot V \rangle}{|\mathrm{N} \cdot \mathrm{V}|} dm$
- $\int_\Omega \frac{F(L,m)\delta_m(H,m)}{4|H \cdot L||H \cdot V|} D(m) G_2(L,V,m) \frac{\langle m \cdot L \rangle}{|N \cdot L|} \frac{\langle m \cdot V \rangle}{|N \cdot V|} dm$
  - $\delta_m(H,m)$ eliminates integral and sets $m = H$
- Specular BRDF: $\frac{F(L,H) D(H) G_2(L,V,H)}{4|N \cdot L||N \cdot V|}$

Microfacet BRDF sub-topic map

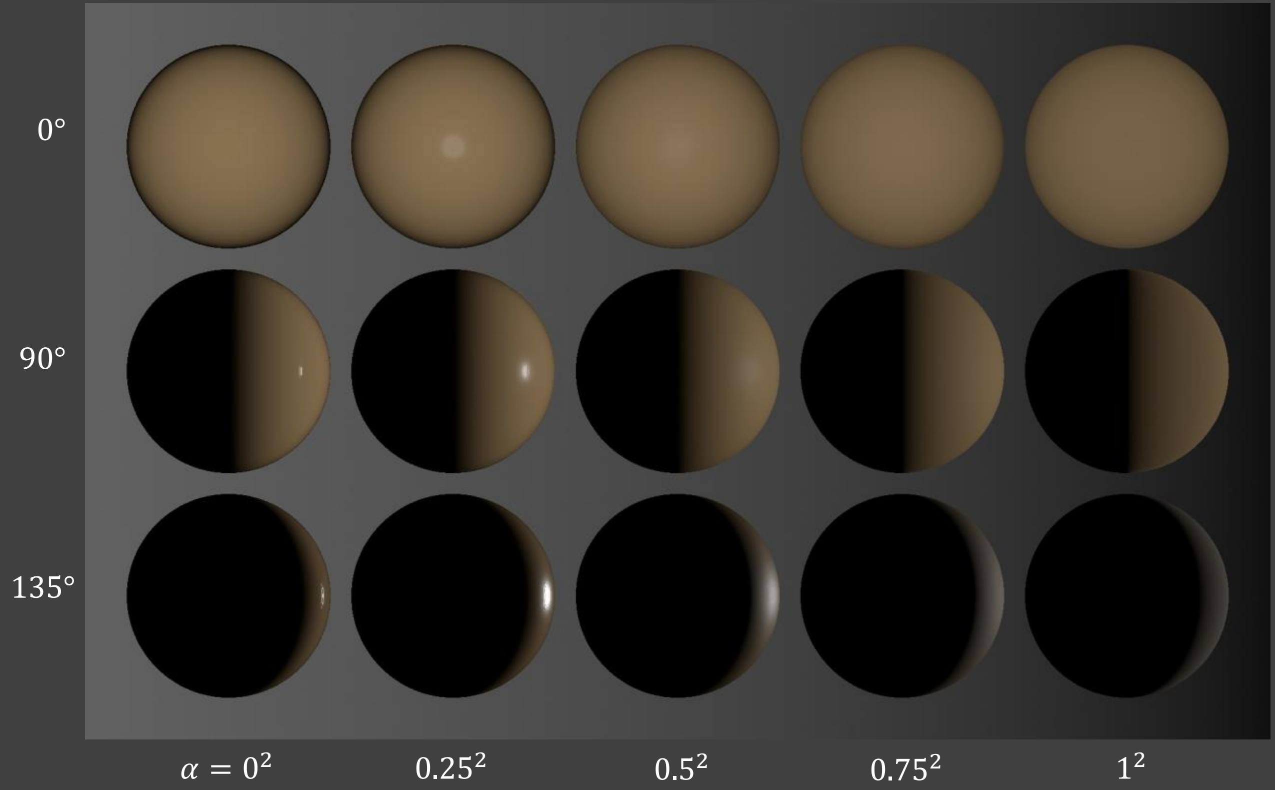
- General form
- How we get PBR specular from that
- Extend to diffuse BRDF

Diffuse Microfacet BRDF

- $\int_\Omega \rho_m(L,V,m) D(m) G_2(L,V,m) \frac{\langle m \cdot L \rangle}{|\mathrm{N} \cdot \mathrm{L}|} \frac{\langle m \cdot V \rangle}{|\mathrm{N} \cdot \mathrm{V}|} dm$
  - Lambertian diffuse: $\rho_m(L,V,m) = \frac{1}{\pi}$
  - No dirac delta to eliminate integral ☹
  - No closed form solution for GGX+Smith ☹☹

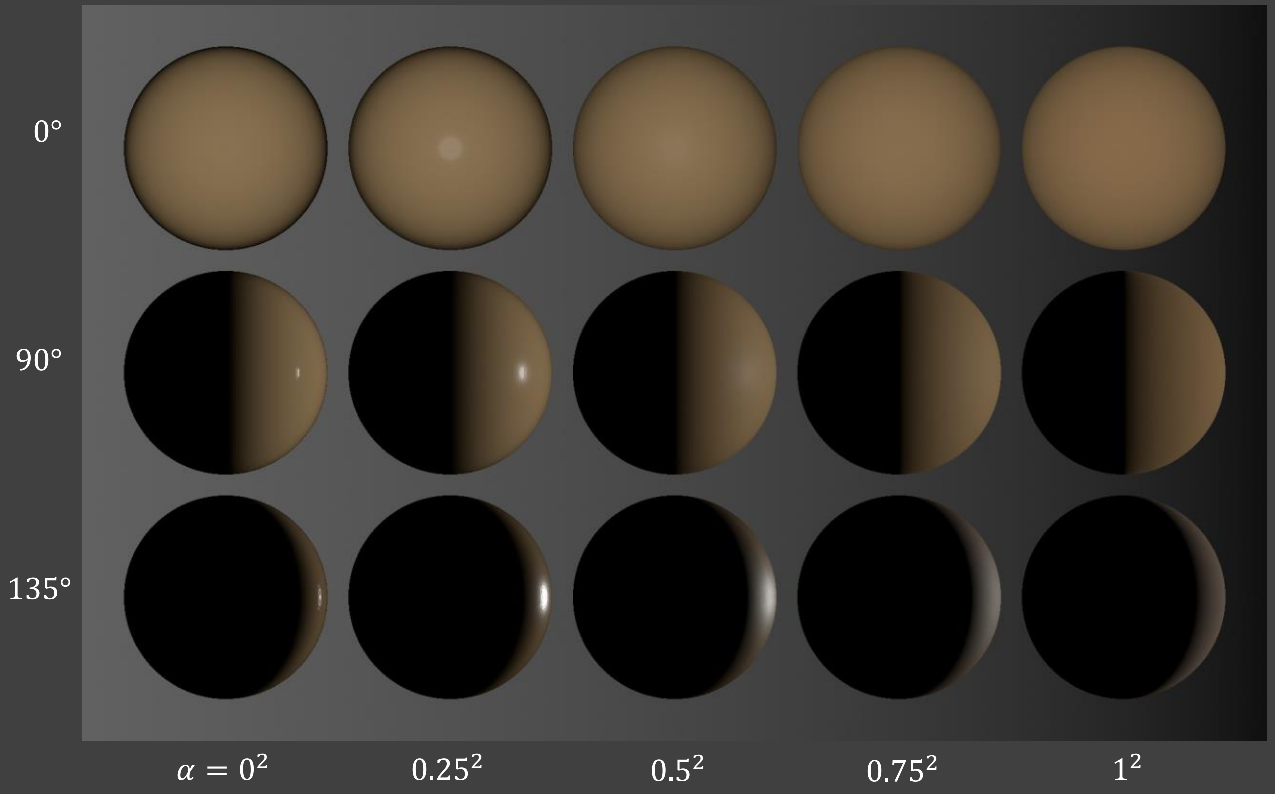
Diffuse Microfacet BRDF

- Solved integral numerically, hoping to find good approximation
    - Same approach as the Oren-Nayar paper
- Up to half the light was missing!
    - Can’t ignore multiple bounces…
        - (Full Oren-Nayar includes a second bounce too)



Direct only

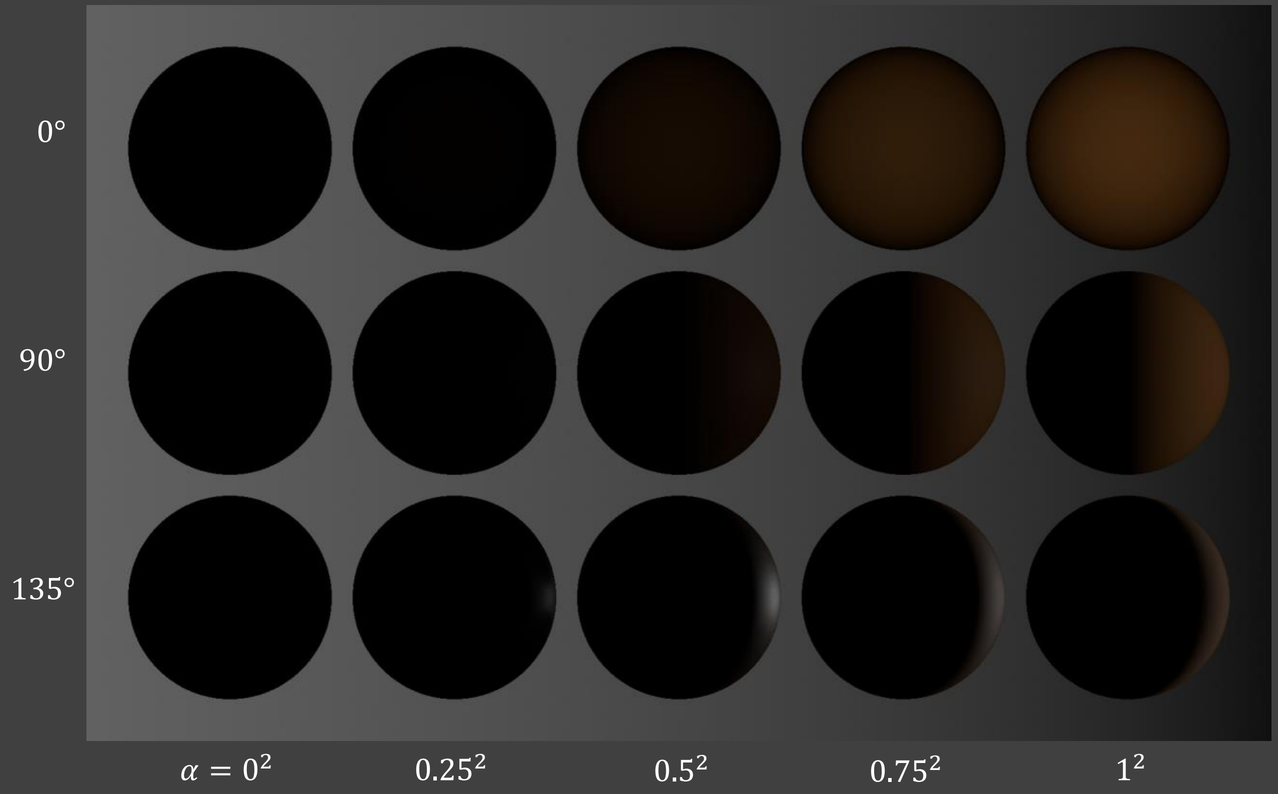
Albedo:
$\{0.75, 0.5, 0.25\}$



Direct plus indirect

Albedo:

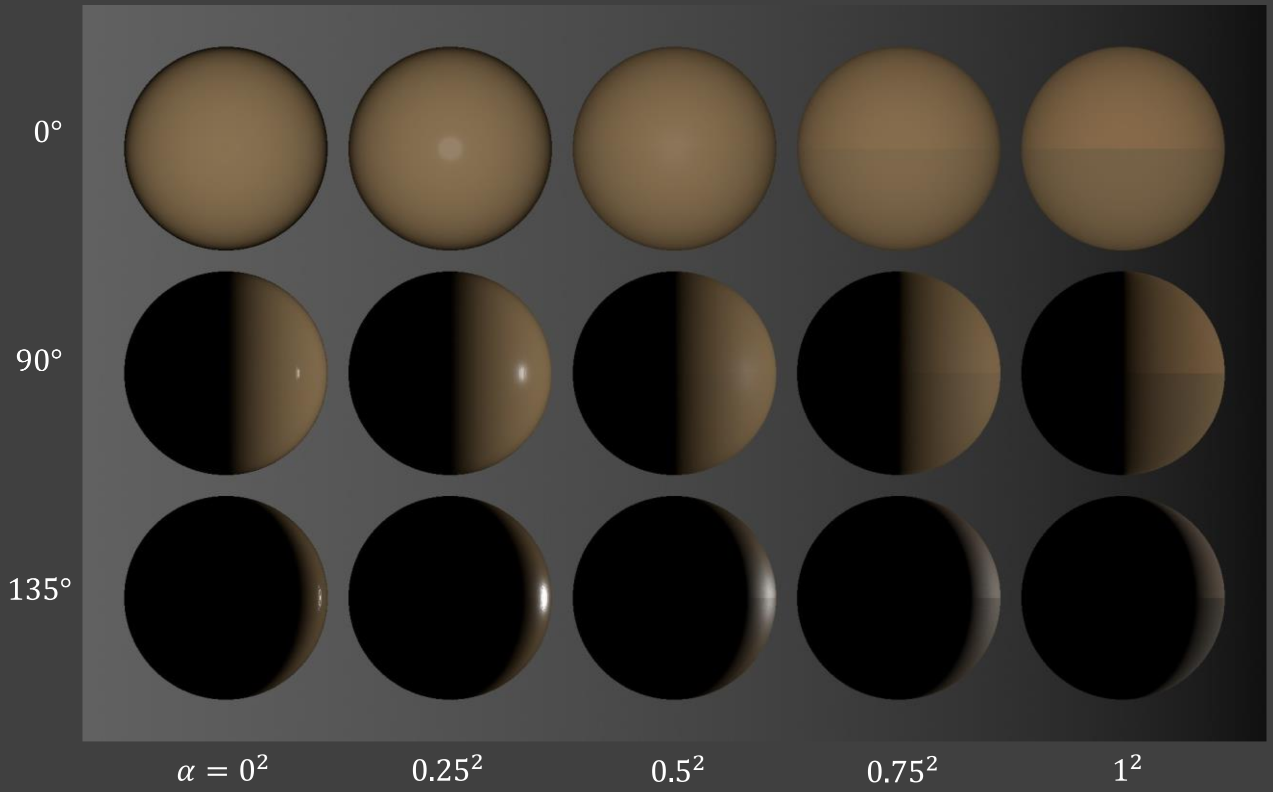
$\{0.75, 0.5, 0.25\}$



Indirect only

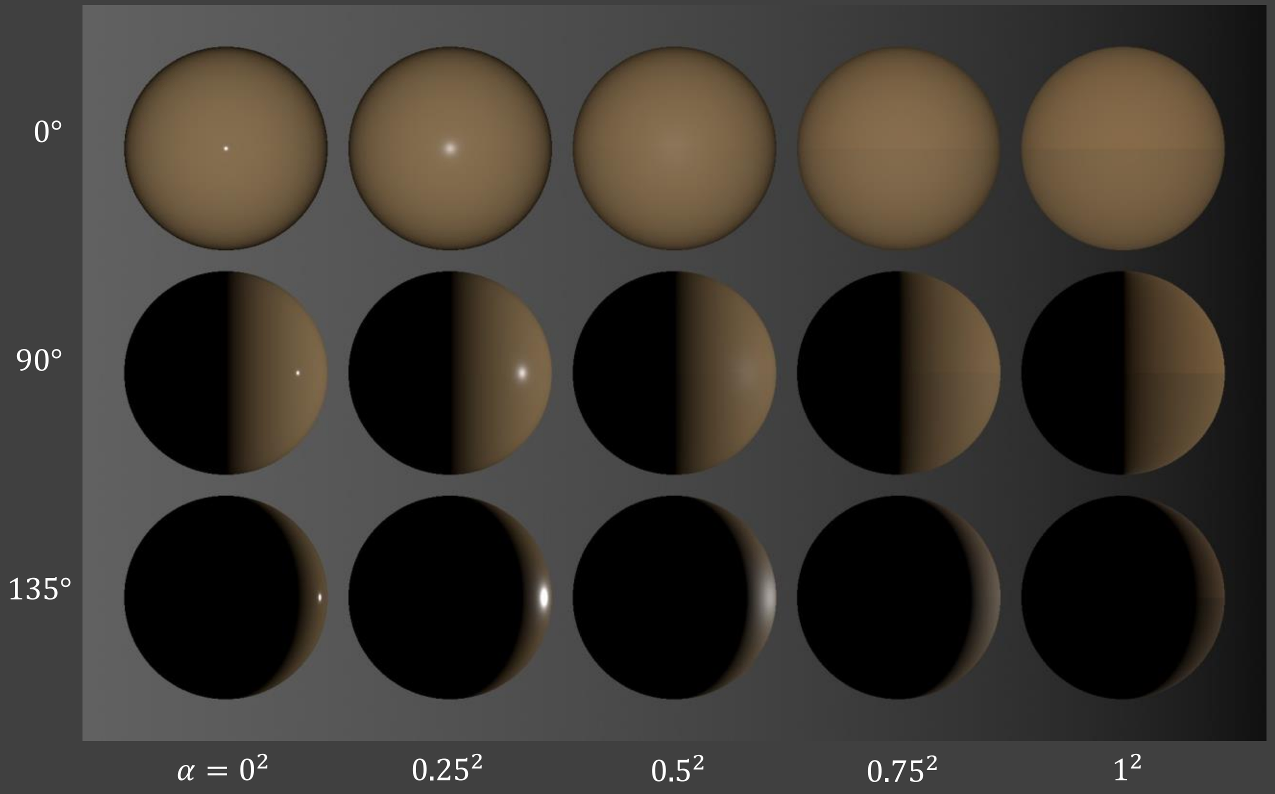
Albedo:
$\{0.75, 0.5, 0.25\}$

Side-by-side



Albedo:

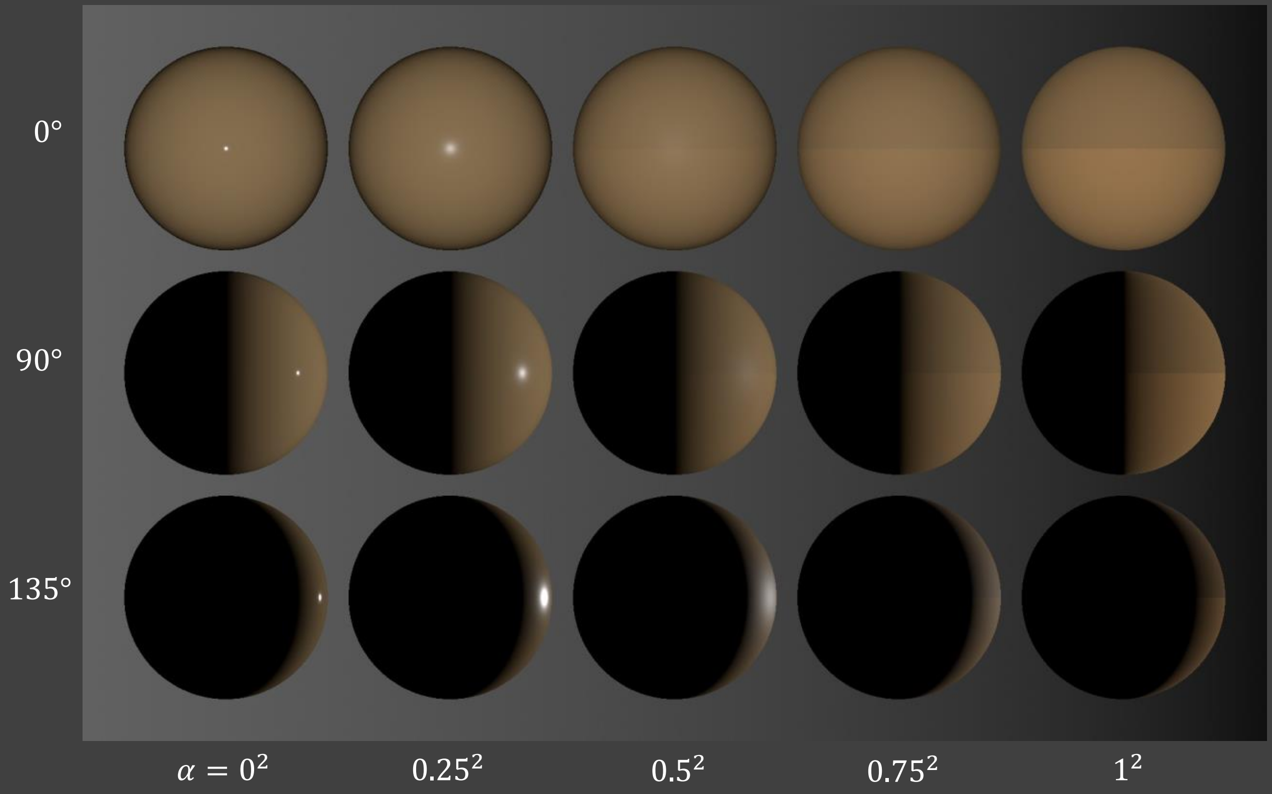
$\{0.75, 0.5, 0.25\}$



Importance of proper albedo

Top:
Correct

Bottom:
Albedo × 0.5,
Light × 2



Importance of proper albedo

Top:
Correct

Bottom:
Albedo × 2,
Light × 0.5

Road map for today’s talk

- General microfacet-based BRDFs
- Simulating diffuse for GGX+Smith microfacet model
- Comparing to other diffuse BRDFs

Road map for today’s talk

- General microfacet-based BRDFs
- Simulating diffuse for GGX+Smith microfacet model
  - Shadowing/masking functions
  - Path tracing
- Comparing to other diffuse BRDFs

Diffuse Simulation sub-topic map

- Shadowing/masking functions (G_1, G_2)
  - Uncorrelated vs height correlated G
  - Smith shadowing/masking
  - New Smith+GGX G_2 approximation
  - Greatness and weirdness of Smith
- Path tracing

Diffuse Simulation sub-topic map

- Shadowing/masking functions (G_1, G_2)
  - Uncorrelated vs height correlated G
  - Smith shadowing/masking
  - New Smith+GGX G_2 approximation
  - Greatness and weirdness of Smith
- Path tracing

Height-Correlated G

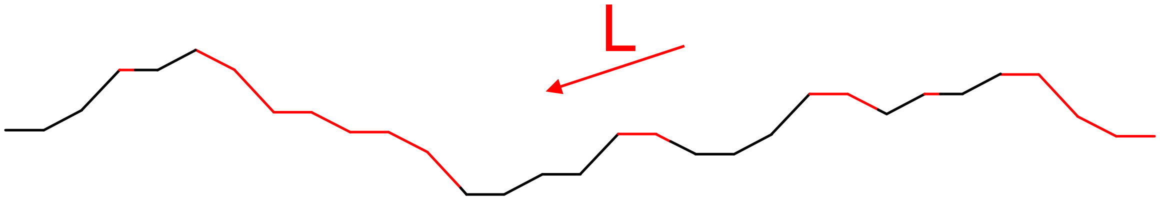
- G_n is geometric visibility to n directions
- If uncorrelated:

$$G_2(L,V,m) = G_1(L,m)G_1(V,m)$$

  - Not realistic! Higher points more likely visible to both L and V (and lower points less likely)
  - Still, surprisingly good in practice

Height-Correlated G

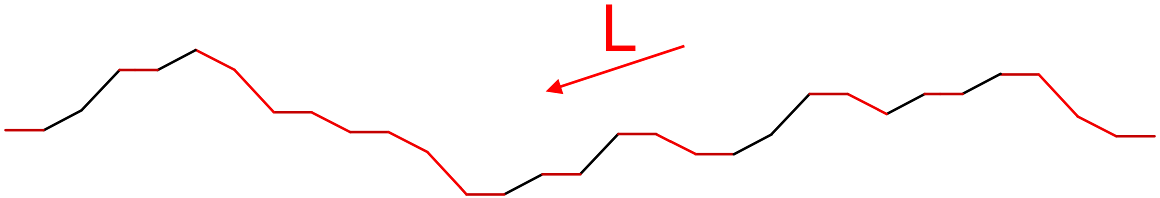
- Uncorrelated G takes light hitting a normal in the heightfield…



m	%
−2	93%
−1	87%
0	57%
+1	0%
+2	0%

Height-Correlated G

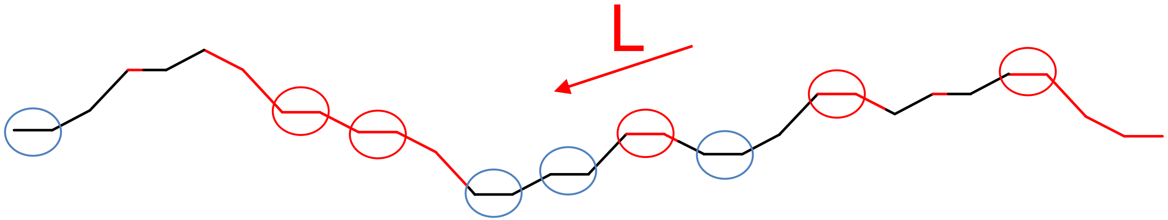
- …and redistributes it evenly across each microfacet with that normal



m	%
-2	93%
-1	87%
0	57%
$+1$	0%
$+2$	0%

Height-Correlated G

- This tends to move light lower, reducing its visibility and darkening specular.

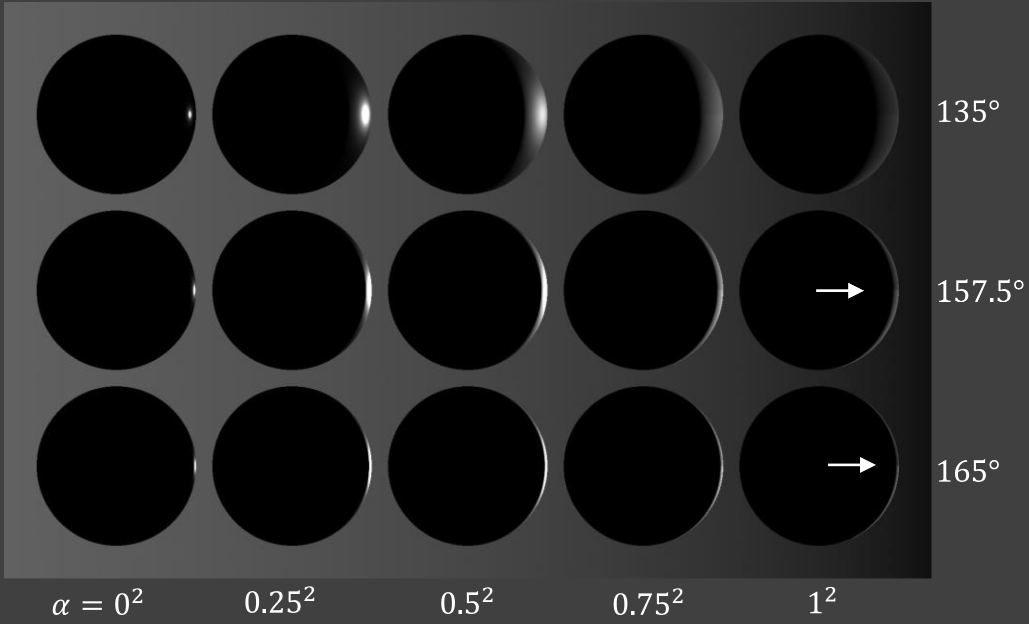


m	%
−2	93%
−1	87%
0	57%
+1	0%
+2	0%

Height-Correlated G

- Uncorrelated G's error is related to occlusion
- Error bigger for rougher surfaces
- Error bigger when L and V more glancing
    - No error if $L = N$ and/or $V = N$

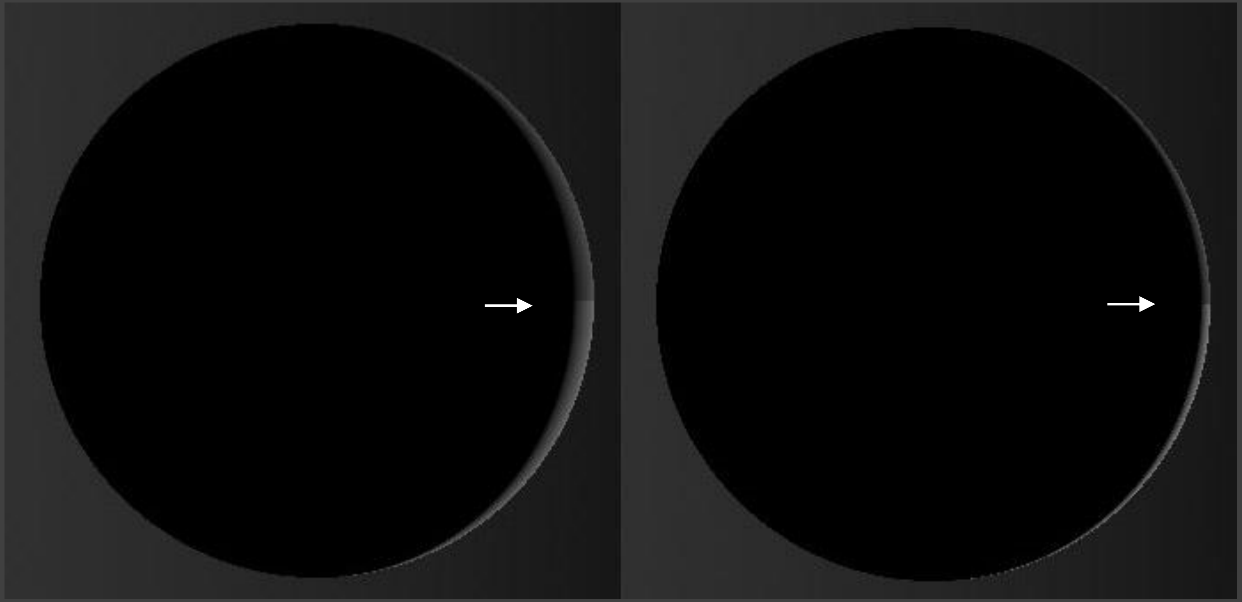
Uncorrelated vs. Correlated G



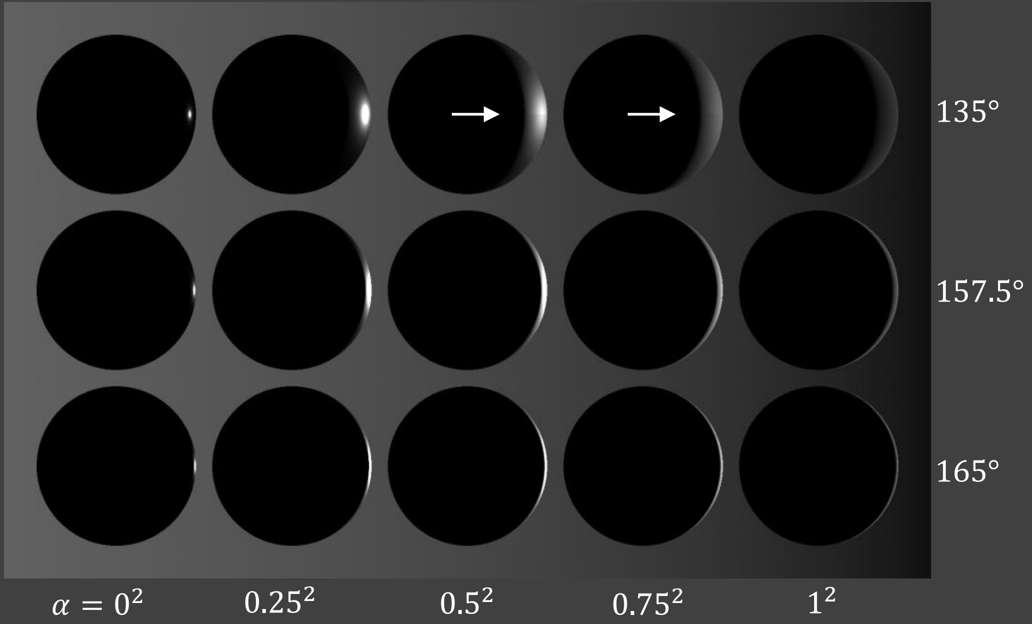
Height-correlation (bottom) boosts glancing reflection on rough surfaces

Black albedo;
light intensity $= \pi$

Uncorrelated vs. Correlated G

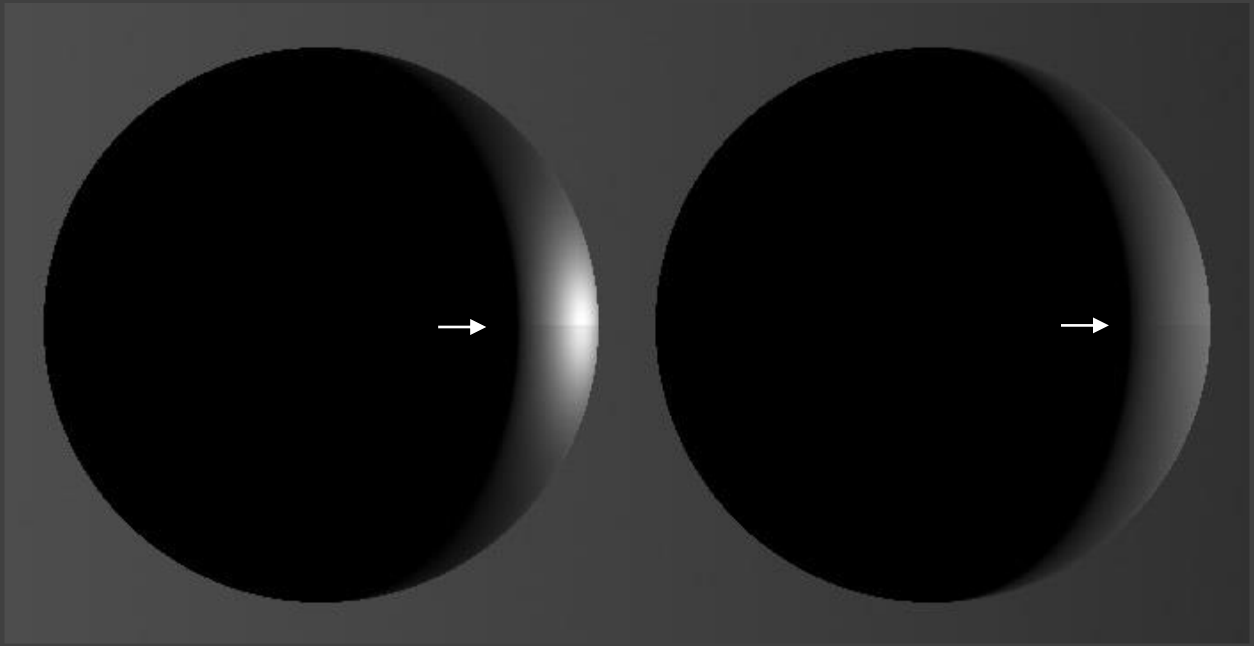


Exact vs. Approx Correlated G



Approximation (bottom) is quite good, but still a little too dark for medium angles and roughness

Exact vs. Approx Correlated G



Height Correlated G

Uncorrelated G

Difference

Correlated G

- There is angular correlation too
- $L = V$ should have: $G_2(V, V, m) = G_1(\mathrm{V}, m)$
- Uncorrelated form: $G_2(V, V, m) = G_1(\mathrm{V}, m)^2$
- Height correlated G_2 somewhere in between when $L = V$

Diffuse Simulation sub-topic map

- Shadowing/masking functions (G_1, G_2)
  - Uncorrelated vs height correlated G
  - Smith shadowing/masking
  - New Smith+GGX G_2 approximation
  - Greatness and weirdness of Smith
- Path tracing

Smith Shadowing/Masking

- Assumes all normals equally occluded
  - I.e., G_1 and G_2 don't depend on m
  - Most balanced assumption possible

Smith Shadowing/Masking

- Can derive from normalization constraint:

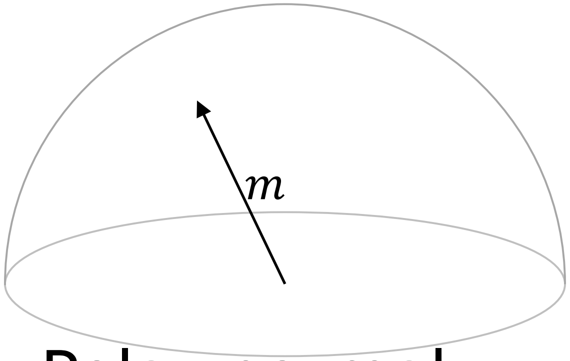
$$G_1(V)\int_\Omega D(m)\langle m\cdot V\rangle dm = |N\cdot V|$$

- Can also derive from ray-tracing a probabilistic heightfield

Smith Shadowing/Masking

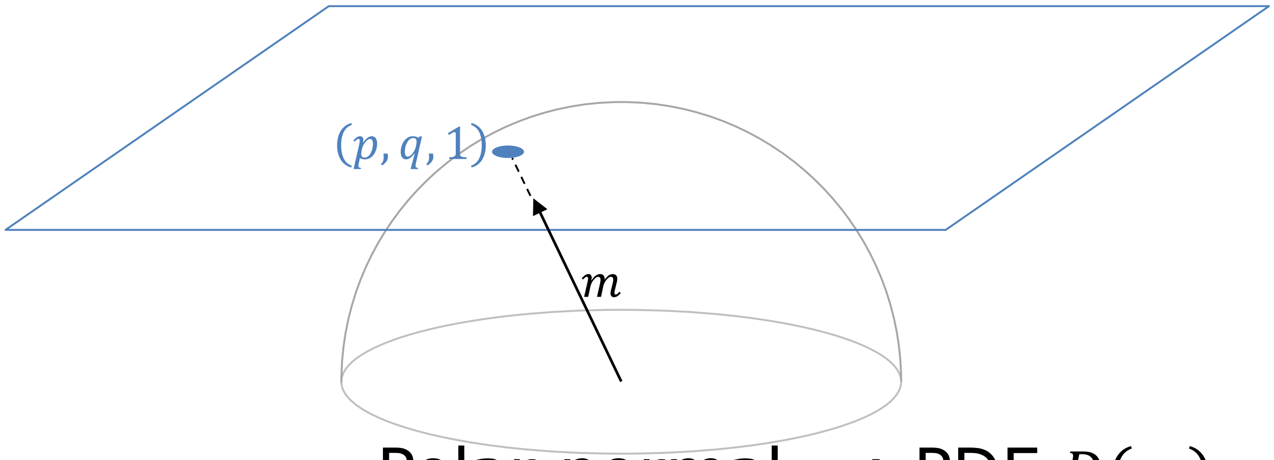
- Super basic ray trace derivation:
  - Project m onto 2D plane with PDF $P_{22}(p,q)$
  - $D(m)$ isotropic, so use 1D slice with PDF $P_2(q)$
  - Use $P_2(q)$ to get PDF of ray-surface collisions while ray with slope μ walks the heightfield
  - Use PDF of collisions to get G_1

Smith Shadowing/Masking



Polar normal m; PDF $D(m)$

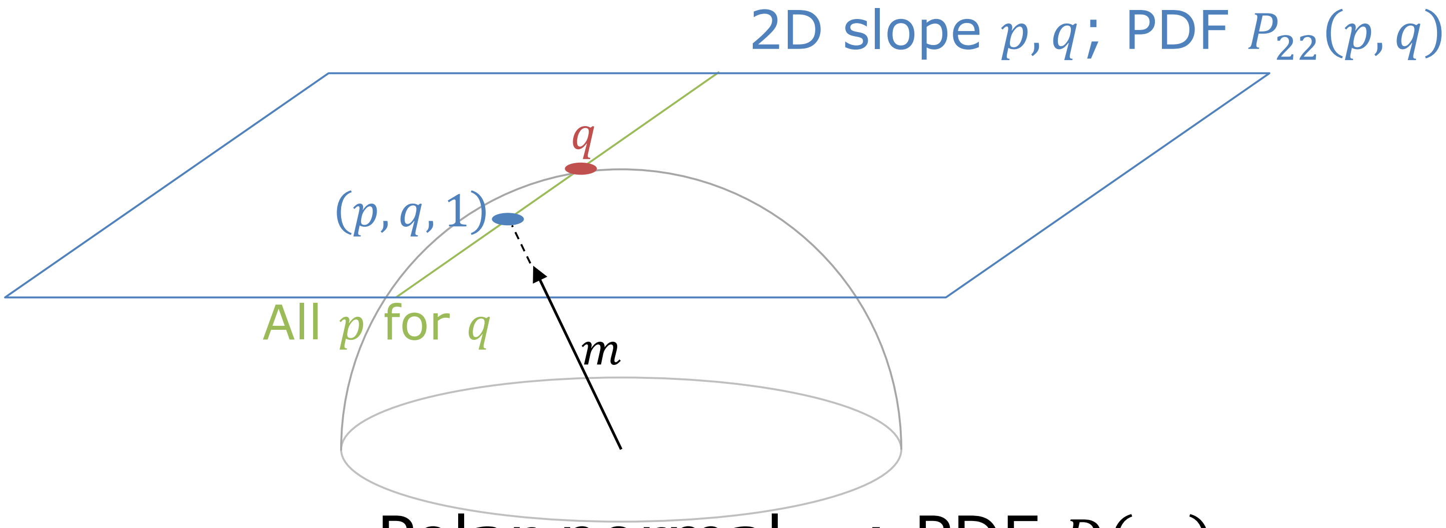
Smith Shadowing/Masking



2D slope p, q; PDF $P_{22}(p, q)$

Polar normal m; PDF $D(m)$

Smith Shadowing/Masking



2D slope p, q; PDF $P_{22}(p, q)$

q

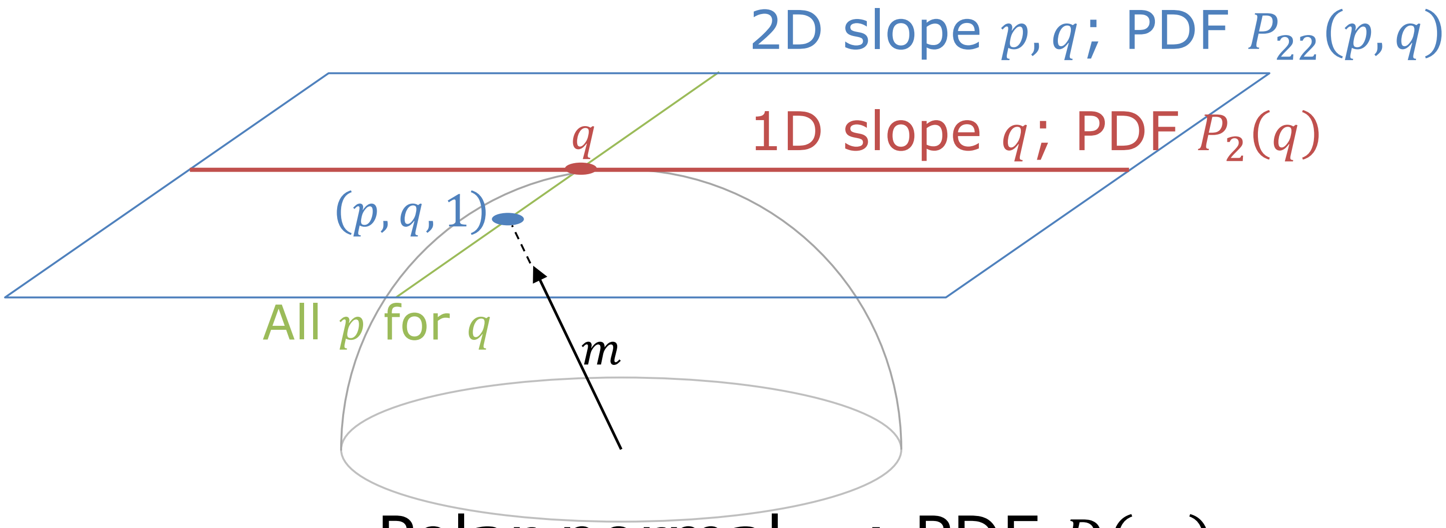
$(p, q, 1)$

All p for q

m

Polar normal m; PDF $D(m)$

Smith Shadowing/Masking



2D slope p, q; PDF $P_{22}(p, q)$

1D slope q; PDF $P_2(q)$

Polar normal m; PDF $D(m)$

Smith: Arbitrary $D(m)$

- $P_{22}(p,q) = \cos^4\theta_m\, D(m)$
- $P_2(q) = \int_{-\infty}^{\infty} P_{22}(p,q)dp$
- $\Lambda(\mu) = \frac{1}{\mu}\int_{\mu}^{\infty}(q-\mu)P_2(q)\,dq$
- $G_1(V) = \frac{1}{1+\Lambda(V)}$ $\qquad G_2(L,V) = \frac{1}{1+\Lambda(L)+\Lambda(V)}$

Smith: Correlated vs Uncorrelated

- $G_1(V) = \frac{1}{1+\Lambda(V)}$
- Correlated: $G_2(L,V) = \frac{1}{1+\Lambda(L)+\Lambda(V)}$
- Uncorrelated: $G_2(L,V) = \frac{1}{1+\Lambda(L)+\Lambda(V)+\Lambda(L)\Lambda(V)}$
  - Too small, unless $\Lambda = 0$ (i.e. $G_1 = 1$) for L or V

Smith for GGX: $\Lambda(V)$

- For GGX: $D(m) = \dfrac{\alpha^2}{\pi(\cos^4\theta_m(\alpha^2 + \tan^2\theta_m)^2)}$
- $P_{22}(p,q) = \dfrac{\alpha^2}{\pi(\alpha^2 + \tan^2\theta_m)^2} = \dfrac{\alpha^2}{\pi(\alpha^2 + p^2 + q^2)^2}$
- $P_2(q) = \int_{-\infty}^{\infty} \dfrac{\alpha^2}{\pi(\alpha^2 + p^2 + q^2)^2} dp = \dfrac{\alpha^2}{2(\alpha^2 + q^2)^{3/2}}$

Smith for GGX: $\Lambda(V)$

- $\Lambda(\mu) = \frac{1}{\mu}\int_{\mu}^{\infty}(\mathrm{q} - \mu)P_2(q)dq = \frac{1}{2}\left(\frac{\sqrt{\alpha^2+\mu^2}}{\mu} - 1\right)$
- $\mu = \cot\theta_V$
- $\cos\theta_V = \mathrm{N}\cdot\mathrm{V}$
- $\Lambda(\mathrm{V}) = \frac{1}{2}\left(\frac{\sqrt{\alpha^2+(1-\alpha^2)(\mathrm{N{\cdot}V})^2}}{\mathrm{N{\cdot}V}} - 1\right)$

Smith for GGX: $G_1(V)$, $G_2(L,V)$

- $G_1(V) = \frac{2\mathrm{N}\cdot\mathrm{V}}{\sqrt{\alpha^2+(1-\alpha^2)(\mathrm{N}\cdot\mathrm{V})^2}+\mathrm{N}\cdot\mathrm{V}}$
- $G_2(L,V) = \frac{2(\mathrm{N}\cdot\mathrm{L})(\mathrm{N}\cdot\mathrm{V})}{\mathrm{N}\cdot\mathrm{V}\sqrt{\alpha^2+(1-\alpha^2)(\mathrm{N}\cdot\mathrm{L})^2}+\mathrm{N}\cdot\mathrm{L}\sqrt{\alpha^2+(1-\alpha^2)(\mathrm{N}\cdot\mathrm{V})^2}}$
  - Would like cheaper approximation!

Diffuse Simulation sub-topic map

- Shadowing/masking functions (G_1, G_2)
  - Uncorrelated vs height correlated G
  - Smith shadowing/masking
  - New Smith+GGX G_2 approximation
  - Greatness and weirdness of Smith
- Path tracing

Smith: Approximate GGX $G_1(V)$

- Denominator of G_1:
  - $\sqrt{\alpha^2 + (1-\alpha^2)(\mathrm{N}\cdot\mathrm{V})^2} + \mathrm{N}\cdot\mathrm{V}$
  - $\sqrt{lerp\left((\mathrm{N}\cdot\mathrm{V})^2, 1, \alpha^2\right)} + \mathrm{N}\cdot\mathrm{V}$
  - Approximation: $lerp(\mathrm{N}\cdot\mathrm{V}, 1, \alpha) + \mathrm{N}\cdot\mathrm{V}$

Smith: Approximate GGX $G_1(V)$

- $G_1(V) \approx \frac{2\mathrm{N}\cdot\mathrm{V}}{\mathrm{lerp}(\mathrm{N}\cdot\mathrm{V},1,\alpha)+\mathrm{N}\cdot\mathrm{V}} = \frac{2\mathrm{N}\cdot\mathrm{V}}{\mathrm{N}\cdot\mathrm{V}(2-\alpha)+\alpha}$
- Turns out, identical to Unreal’s Smith:
  - $G_1(V) \approx \frac{\mathrm{N}\cdot\mathrm{V}}{\mathrm{N}\cdot\mathrm{V}(1-k)+k}, k = \frac{\alpha}{2}$

Smith: Approximate GGX $G_2(L,V)$

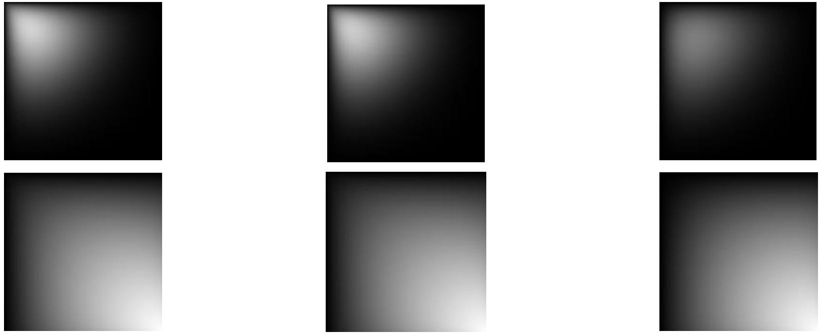
- Solve this G_1 for $\Lambda(V)$, plug in for $G_2(L,V)$:
  - $G_2(L,V) = \dfrac{2|N\cdot L||N\cdot V|}{\text{lerp}(2|N\cdot L||N\cdot V|, |N\cdot L| + |N\cdot V|, \alpha)}$
  - G_2's numerator cancels in full specular BRDF:
    - $\dfrac{F(L,H)D(H)G_2(L,V)}{4|N\cdot L||N\cdot V|} = \dfrac{F(L,H)D(H)}{2\ \text{lerp}(2|N\cdot L||N\cdot V|, |N\cdot L| + |N\cdot V|, \alpha)}$

Smith Approximation Cost

- Compare cost of denominator:
  - $G_1(L)G_1(V)$: $\frac{F(L,H)D(H)}{(|N\cdot L|(2-\alpha)+\alpha)(|N\cdot V|(2-\alpha)+\alpha)}$ ~4 cycles
  - $G_2(L,V)$: $\frac{F(L,H)D(H)}{2\,\mathrm{lerp}(2|N\cdot L||N\cdot V|,|N\cdot L|+|N\cdot V|,\alpha)}$ ~6 cycles
    - Costs exclude calculating $N\cdot L$ and $N\cdot V$
    - Height-correlated form has negligible extra cost

Smith Approximation Quality

- Helps rough dielectrics at glancing angles



Exact $G_2(L,V)$ Approx $G_2(L,V)$ Approx $G_1(L)G_1(V)$

$F_0 = 0, I = 20$

$F_0 = 1, I = 2$

Top left of image: glancing reflection

Bottom right: normal incidence and viewer

GGX Specular BRDF for $\alpha = 0.8$; $N \cdot L, N \cdot V$ increase down,right

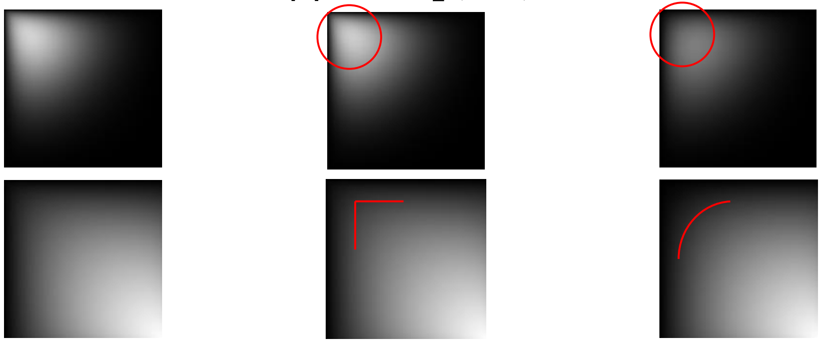
Smith Approximation Quality

- Helps rough dielectrics at glancing angles

Exact $G_2(L,V)$ Approx $G_2(L,V)$ Approx $G_1(L)G_1(V)$

$F_0 = 0, I = 20$

$F_0 = 1, I = 2$



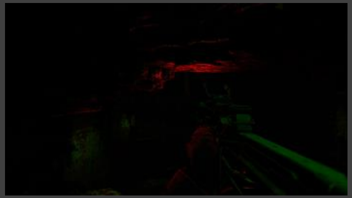
Top left of image: glancing reflection

Bottom right: normal incidence and viewer

GGX Specular BRDF for $\alpha = 0.8$; $N \cdot L, N \cdot V$ increase down,right

Uncorrelated G Approximation

Difference Image:



Red = correlation
Green = approximation

Correlated G Approximation

Difference Image:

Relative to uncorrelated approximation

Correlated G Exact

Difference Image:

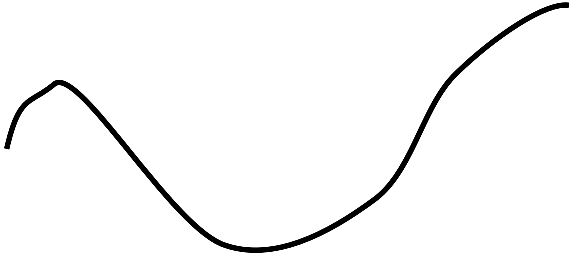
Relative to correlated approximation

Diffuse Simulation sub-topic map

- Shadowing/masking functions (G_1, G_2)
  - Uncorrelated vs height correlated G
  - Smith shadowing/masking
  - New Smith+GGX G_2 approximation
  - Greatness and weirdness of Smith
- Path tracing

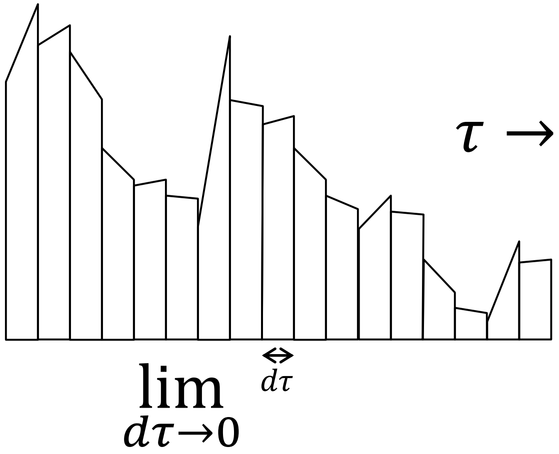
Smith Microsurface Heightfields

- Example 1D heightfield from Smith ray tracing derivation sketch in Walter 2007



Smith Microsurface Heightfields

- Derivation *also* uses 1D heightfield of mostly independent slabs nearing zero width
  - Only forbids suddenly being under heightfield



Why Smith masking is weird

- Ray tracing derivation has contradictory assumptions at different steps:
  - Height in next $d\tau$ independent of this height
    - Assumes not continuous
  - Heightfield is any differentiable function
    - Assumes continuous

Why Smith masking is weird

- Math says visibility is asymmetric: downward rays less likely than upward rays to survive the same heightfield path!
  - $\Lambda(\mu) = \frac{1}{\mu}\int_{\mu}^{\infty}(q-\mu)P_2(q)\,dq$
  - $\Lambda(\mu)$ integrates all $q > \mu$, so $\mu < 0$ can hit more values of q than when $\mu > 0$

Why Smith masking is great

- Only energy conserving G where all facet normals have the same fraction visible
- Any other G not using m gets total visible area wrong for some directions
  - Too high reflects too much, creating energy
  - Too low reflects too little, absorbing energy

Diffuse Simulation sub-topic map

- Shadowing/masking functions (G_1, G_2)
  - Uncorrelated vs height correlated G
  - Smith shadowing/masking
  - New Smith+GGX G_2 approximation
  - Greatness and weirdness of Smith
- Path tracing

Path traced diffuse solution

- Smith oddities prevent real heightfields from matching its assumptions
  - Can’t be both continuous and discontuous
- Must ray trace the mathematical model
  - See bonus slides for numerous details

First ray traced result

- Simple ray tracer with Fresnel to choose GGX specular or Lambertian diffuse
- Resulting BRDF was not symmetric!
  - $\rho(L, V, N) \neq \rho(V, L, N)$
- What went wrong? Both parts are symmetric BRDFs!

Cause of asymmetric BRDF

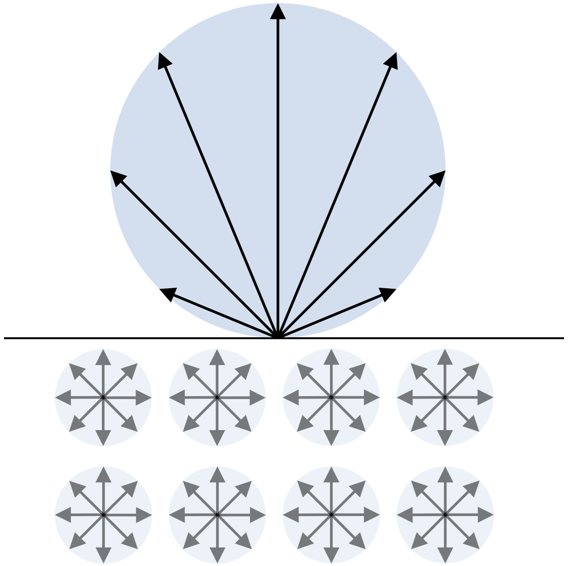
- Essentially had a merged BRDF:
  - $\rho = F(L,N)\rho_{spec} + \left(1 - F(L,N)\right)\rho_{diff}$
  - Fresnel interpolation asymmetric!
- How to fix in a physically plausible way?

Why Lambertian diffuse?

- What does Lambertian diffuse simulate?
  - BRDF $\rho = \frac{1}{\pi}$: same for all viewers
  - Radiance $= \rho \cos \theta_V$: more photons at normal
    - Balanced by $\frac{1}{\cos \theta_V}$ for total surface area seen by V
    - Why the cosine energy falloff? Answer is surprisingly hard to discover, yet quite simple!

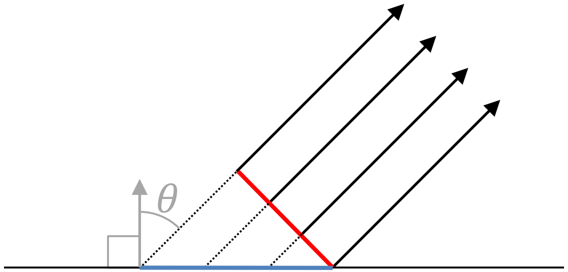
Lambertian diffuse explained

- BRDFs given at the surface, but diffuse light just passes through the surface
  - Lambert assumes interior light has same density in all directions
  - Cosine falloff is from surface angle relative to light direction



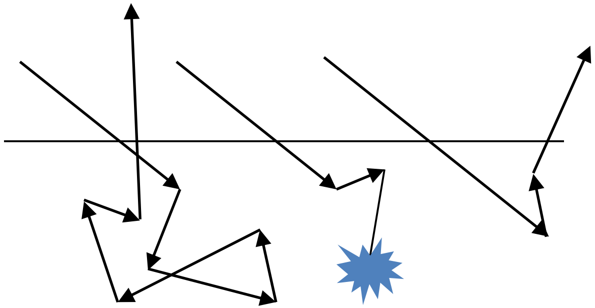
Lambertian diffuse explained

- Same energy per area each direction
- Directions angled to surface project over larger area
  - Area per unit light scaled by $^1/_{\cos\theta}$
  - Light per unit area scaled by $\cos\theta$



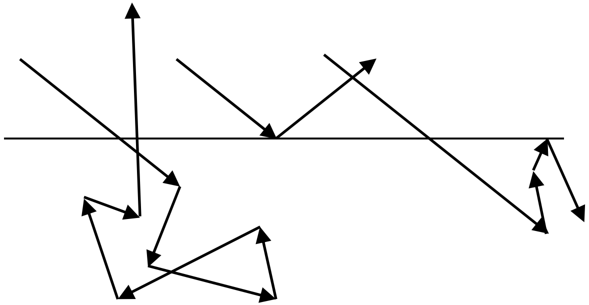
Lambertian diffuse explained

- Light enters, bounces around, exits
- Exit direction is random after many bounce events
- Albedo comes from frequency-dependent absorption events



Fixing diffuse for symmetric BRDF

- Entering uses Fresnel for reflection/transmission
- Exiting assumes always transmit
- Exiting needs Fresnel too!



Fixing diffuse for symmetric BRDF

- Reflect: $F = F_0 + (1 - F_0)(1 - N \cdot V)^5$
- Transmit: $1 - F = (1 - F_0)(1 - (1 - N \cdot V)^5\ )$
- Fresnel’s laws are symmetric, so fraction entering surface from viewer equals fraction exiting surface toward viewer

Fixing diffuse for symmetric BRDF

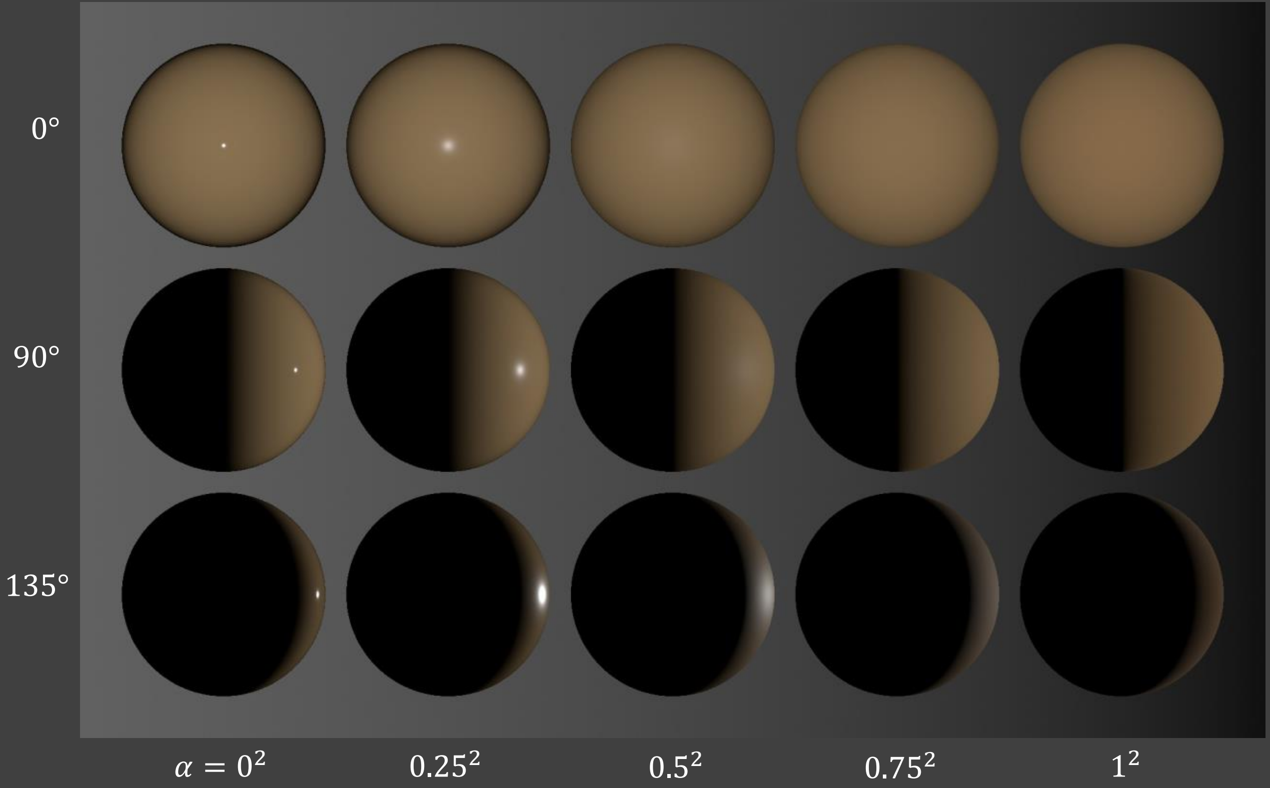
- Internally reflected light keeps getting chances to transmit; need to normalize!
  - $2\pi \int_0^{\pi/2} k(1-(1-\cos\theta)^5) \cos\theta \sin\theta \, d\theta = 1$
    - Factor $(1-F_0)$ absorbed into norm factor k
    - $\cos\theta$ needed to normalize a BRDF
    - 2π and $\sin\theta \, d\theta$ from integrating on a hemisphere

Fixing diffuse for symmetric BRDF

- Easily solved exactly: $k = \frac{21}{20\pi} = \frac{1.05}{\pi}$
- Merged diffuse+spec microfacet BRDF:
  - $F = F_0 + (1 - F_0)(1 - N \cdot V)^5$
  - $\rho = F\,\rho_{spec} + (1 - F)\frac{1.05}{\pi}(1 - (1 - N \cdot V)^5)$

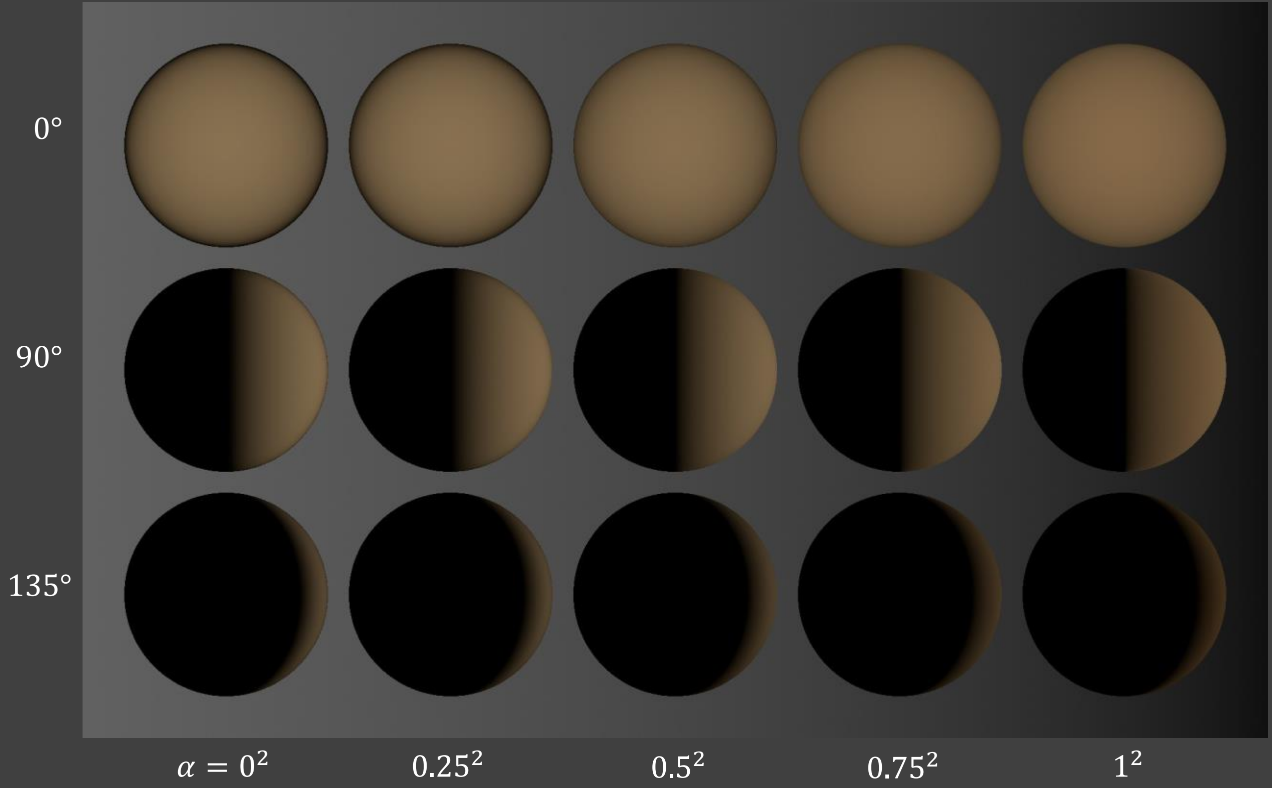
Finally!

- Now have everything needed for path tracing simulation, resulting in...



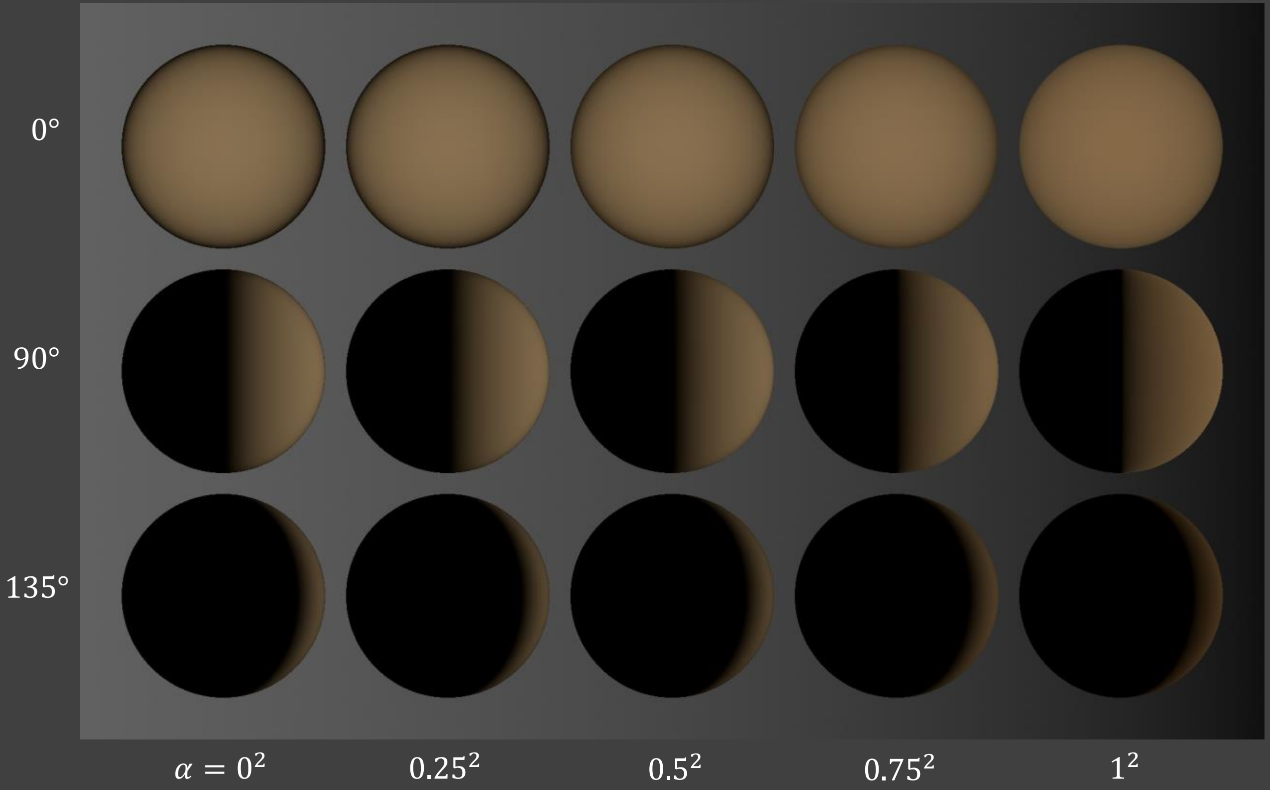
Simulation with spec

Albedo: $\{0.75, 0.5, 0.25\}$



Simulation diffuse only

Albedo: $\{0.75, 0.5, 0.25\}$



Approximate diffuse only

Albedo:
$\{0.75, 0.5, 0.25\}$

GGX Diffuse Approximation

- $facing = 0.5 + 0.5\, L \cdot V$
- $rough = facing(0.9 - 0.4 facing)\left(\frac{0.5+N\cdot H}{N\cdot H}\right)$
- $smooth = 1.05(1 - (1 - N \cdot L)^5)(1 - (1 - N \cdot V)^5)$
- $single = \frac{1}{\pi} lerp(smooth, rough, \alpha)$
- $multi = 0.1159\alpha$
- $diffuse = albedo * (single + albedo * multi)$

Aside: Useful shader identities

- $|L+V|^2 = 2 + 2L \cdot V$
  - $0.5 + 0.5L \cdot V = \frac{1}{4}|L+V|^2$
- $N \cdot H = \frac{N \cdot L + N \cdot V}{|L+V|}$
- $L \cdot H = V \cdot H = \frac{1}{2}|L+V|$

Aside: Useful shader identities

- Can find $N \cdot H$ and $L \cdot H$ without finding H!

Calculation	Cycles	Registers
Get $H = \mathrm{normalize}(L + V)$	13	4
Get H then $N \cdot H$	16	4
Get H then $N \cdot H$ and $L \cdot H$	19	4
$\boldsymbol{N \cdot H}$ from identities	7*	2
$\boldsymbol{L \cdot H}$ and $\boldsymbol{V \cdot H}$ from identities	8*	2

* Add 3 cycles if you don't already have $L \cdot V$

Aside: Useful shader identities

- $lenSq\_LV = 2 + 2\, L \cdot V$
- $rcpLen\_LV = rsqrt(lenSq\_LV)$
- $N \cdot H = (N \cdot L + N \cdot V) * rcpLen\_LV$
- $L \cdot H = V \cdot H = rcpLen_{LV} + rcpLen_{LV} * L \cdot V$
    - (Since $L \cdot H = \frac{1}{2}|L + V| = \frac{1}{2}\sqrt{2 + 2L \cdot V} = \frac{1}{2}\left(\frac{2+2L\cdot V}{\sqrt{2+2L\cdot V}}\right) = (1 + L \cdot V)\frac{1}{\sqrt{2+2L\cdot V}}$)

Road map for today’s talk

- General microfacet-based BRDFs
- Simulating diffuse for GGX+Smith microfacet model
  - Shadowing/masking functions
  - Path tracing
- Comparing to other diffuse BRDFs

But First…

- It’s good to quickly understand Disney’s BRDF slices

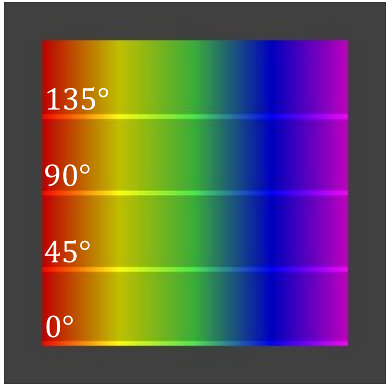
Disney’s BRDF slices

- BRDF is a 4D function of 2 polar vectors
- Before, light+viewer vectors: $\theta_l, \phi_l, \theta_v, \phi_v$
- After, half angle+difference: $\theta_h, \phi_h, \theta_d, \phi_d$
  - Isotropic BRDFs <u>never</u> depend on ϕ_h
  - Dependence on ϕ_d is often negligible

Disney’s BRDF slices intuition

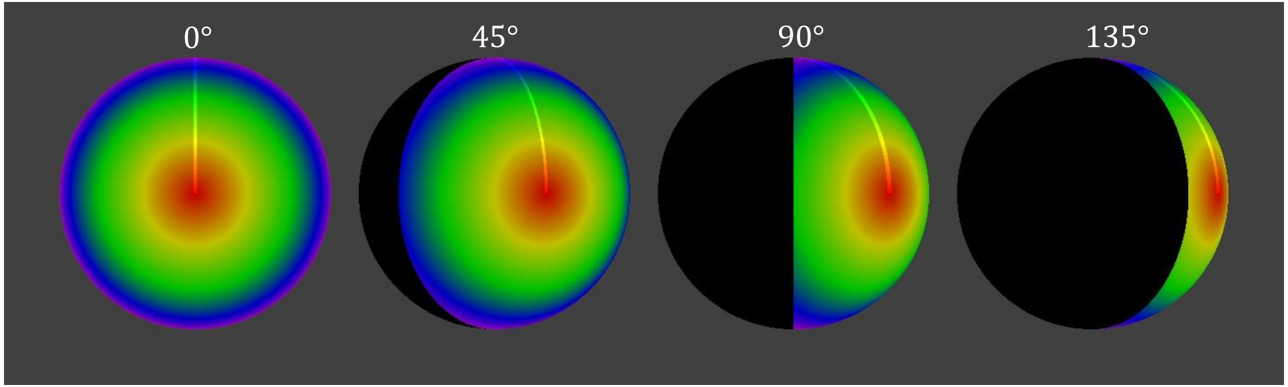
- Each row is a light+viewer pair (θ_d)
  - Opposite at top, perpendicular in middle, coincident at bottom
- Left-to-right shows falloff going away from center of specular highlight (θ_h)

False color example on lit sphere



BRDF Slice

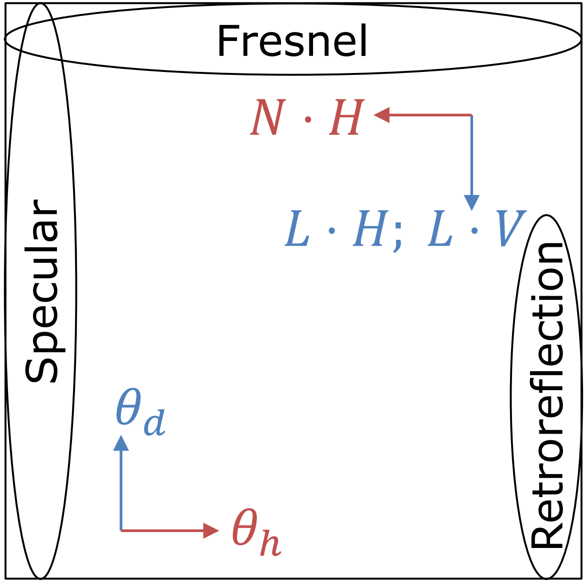
Lighter bands highlight rows used by spheres

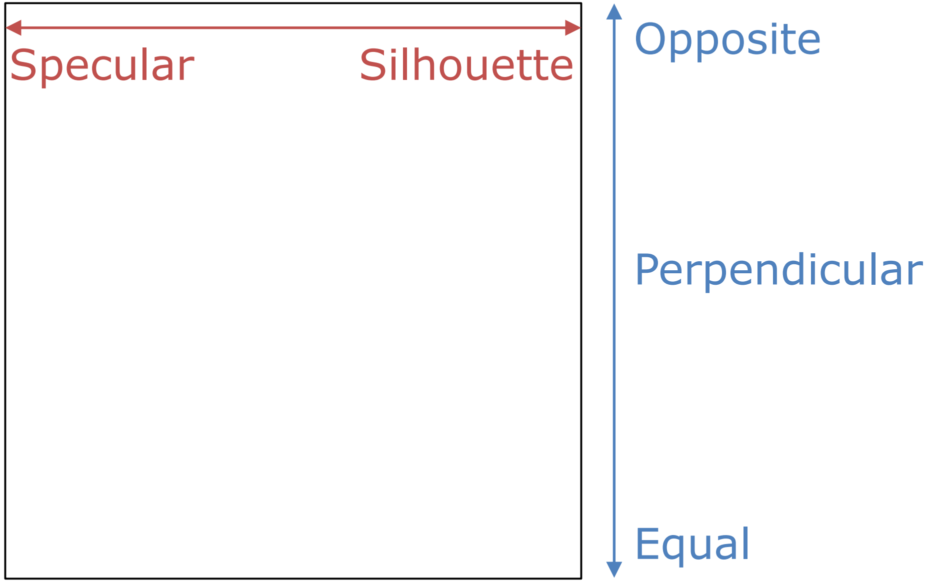


Corresponding Lit Spheres

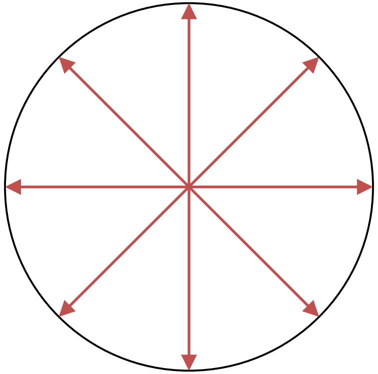
Lighter bands highlight $\phi_d = 90°$; ϕ_d increases counterclockwise

Disney’s BRDF slices

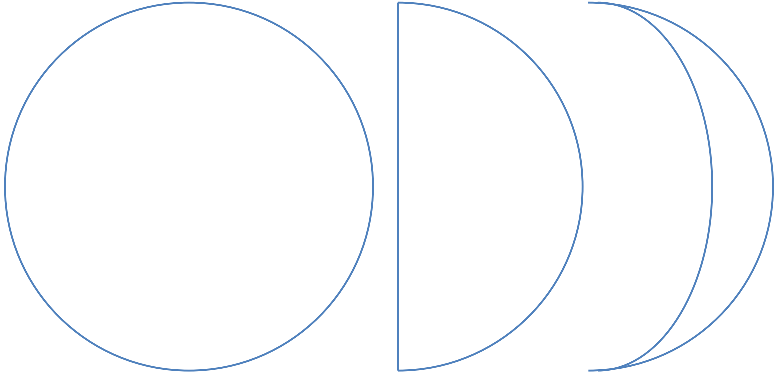




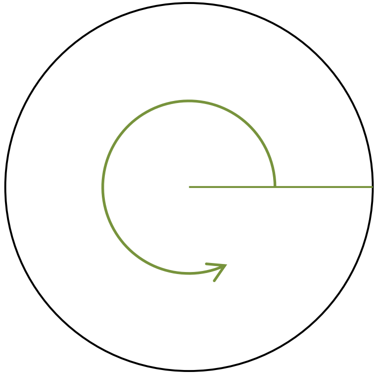
Behavior of $\theta_h, \theta_d, \phi_d$ on spheres



$\theta_h; N \cdot H$



$\theta_d; L \cdot H; L \cdot V$



ϕ_d

Disney’s BRDF slices

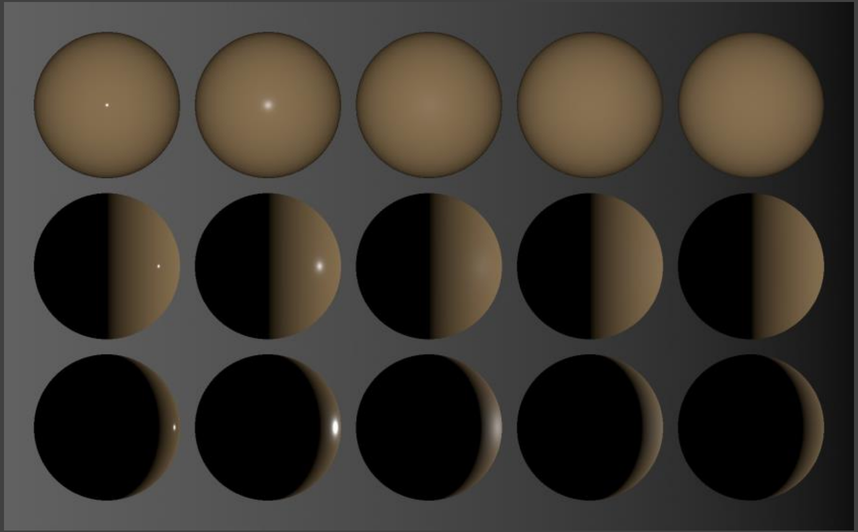
- Various identities:
  - $\cos\theta_h = N \cdot H$
  - $\cos\theta_d = L \cdot H = V \cdot H$ $\cos 2\theta_d = L \cdot V$
  - $\cos\phi_d = \frac{N \cdot V - N \cdot L}{\sqrt{(2-2L \cdot V)(1-(N \cdot H)^2)}}$
- BRDFs mostly functions of $N \cdot H$ and $L \cdot V$!

Almost ready to compare BRDFs!

- First, introduce the comparison format

New Model

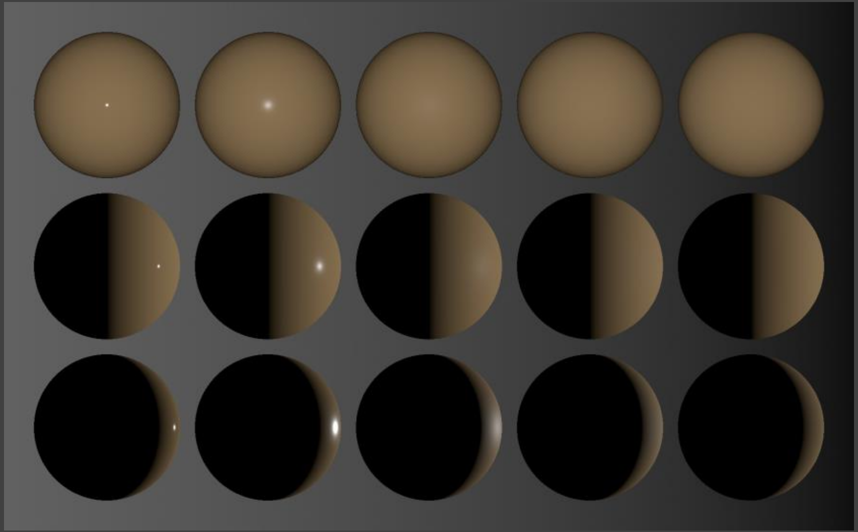
Title says which diffuse model is shown. This intro uses the new model

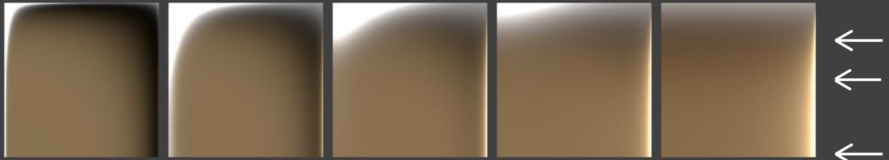


New Model

← This panel shows the same lit spheres as previous examples.

New Model

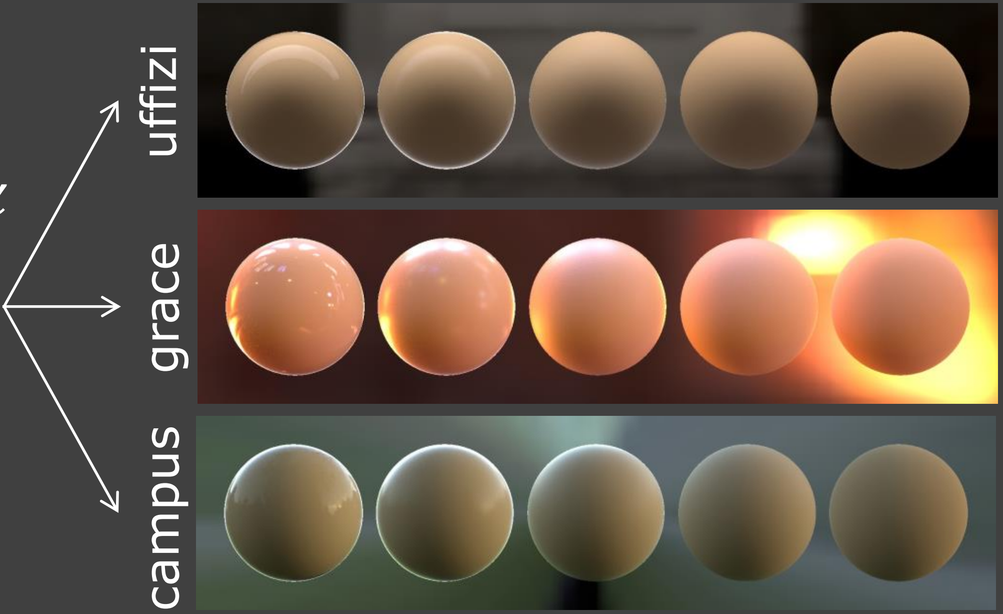




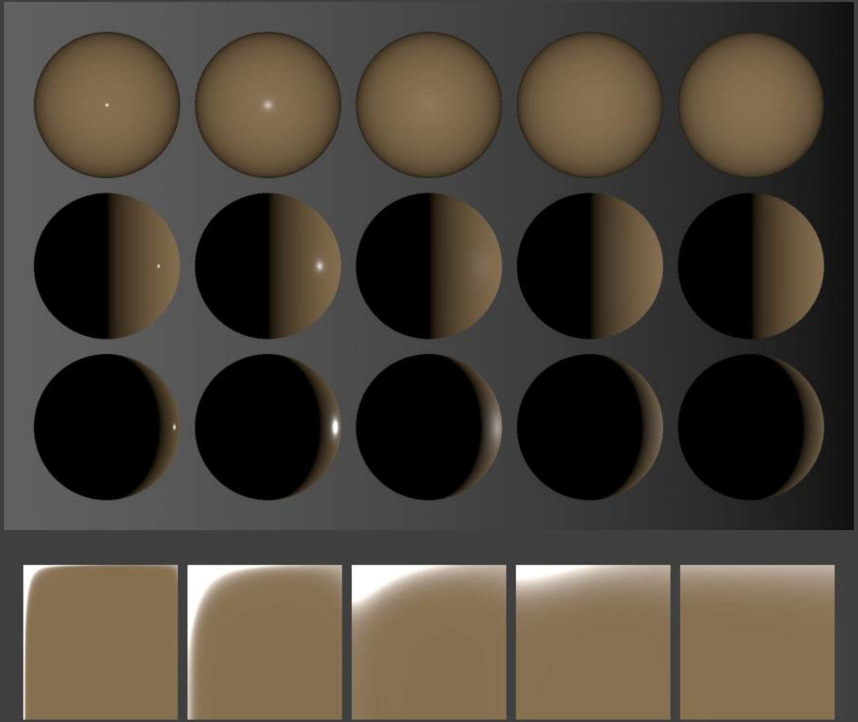
The matching BRDF slices are here (uncorrelated G)

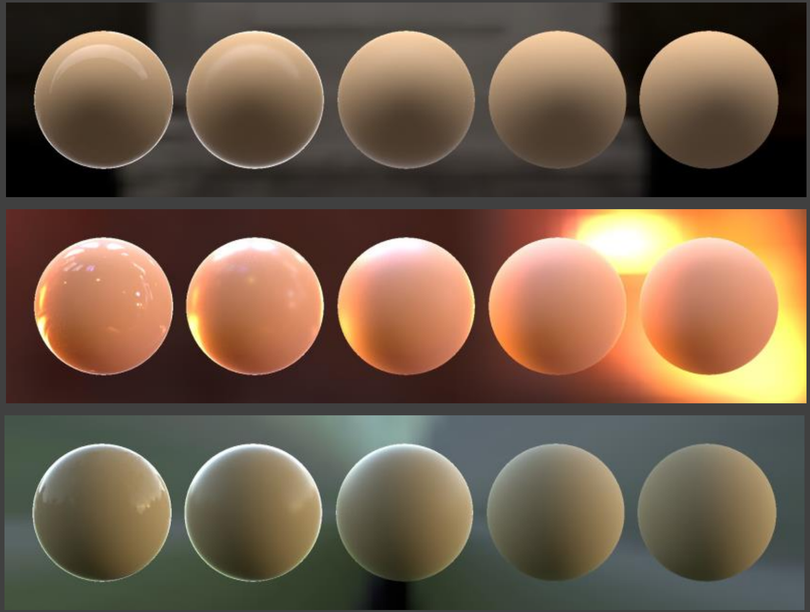
New Model

Same full BRDF with α from 0 to 1, but lit by Paul Debevec’s HDR environment probes.

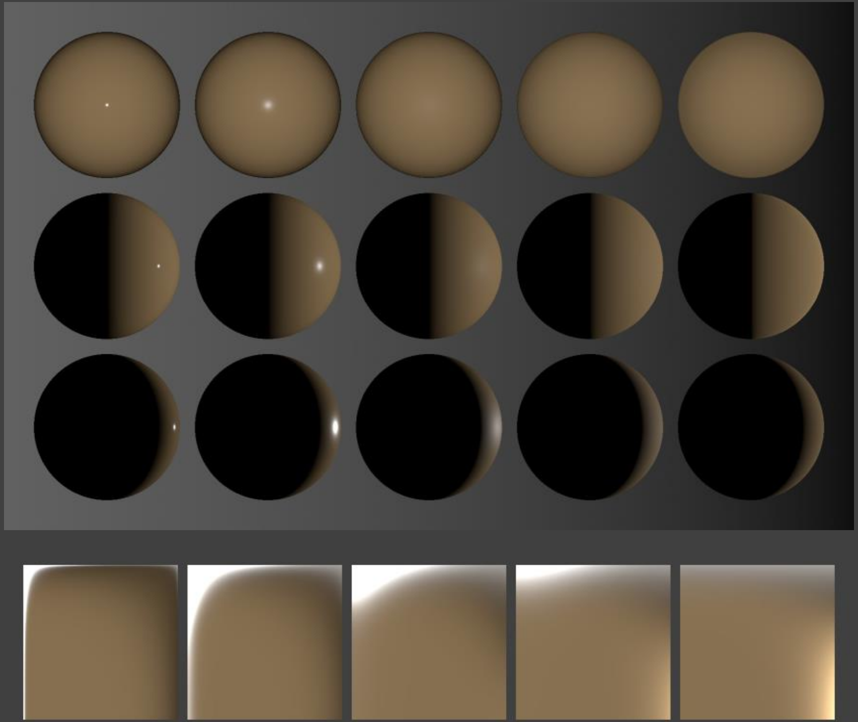


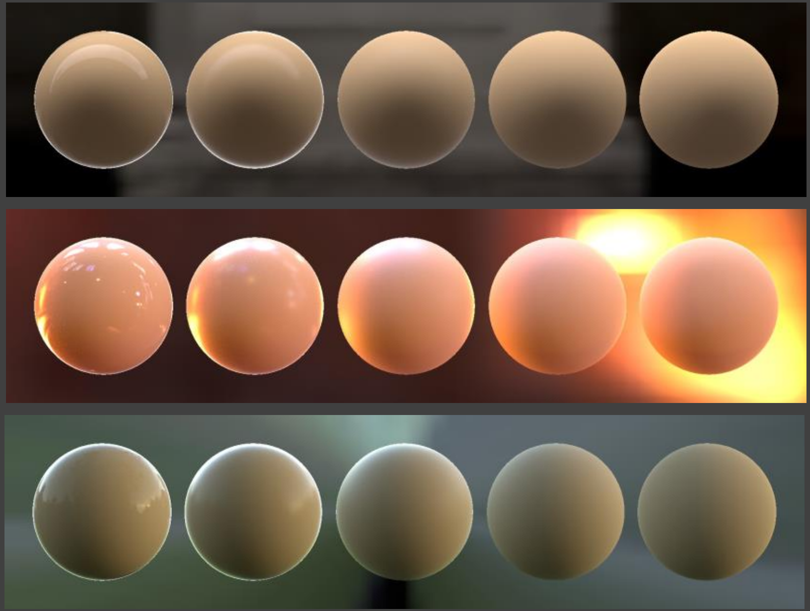
Lambert



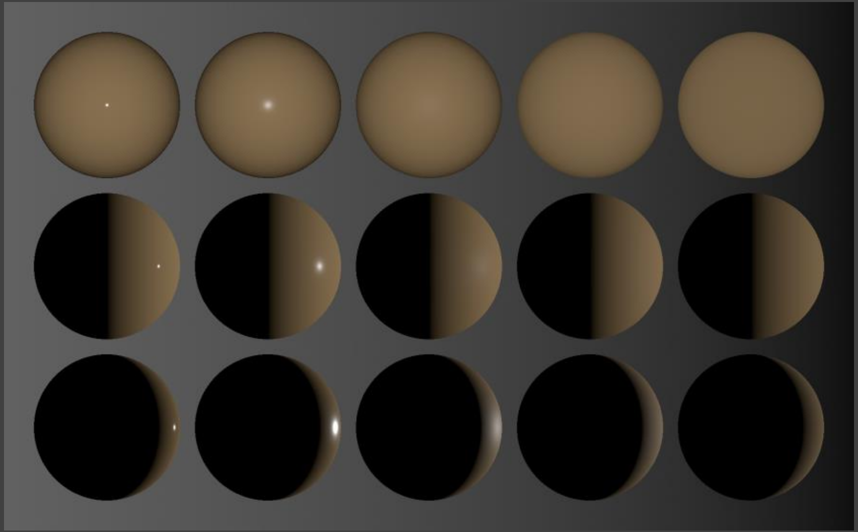


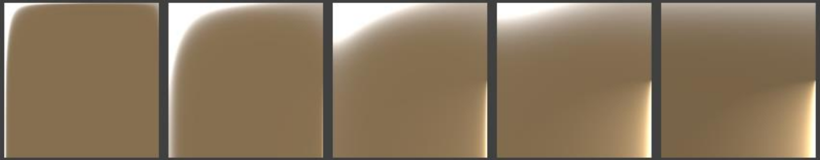
Disney

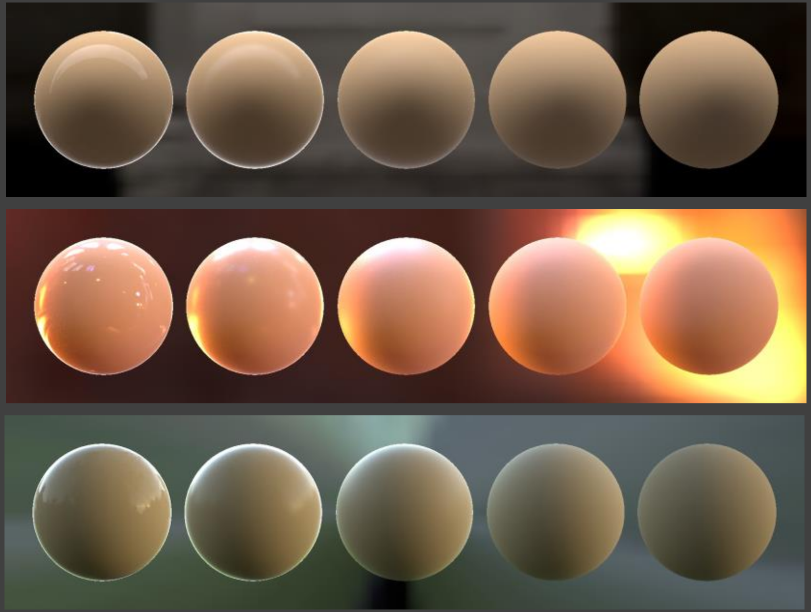




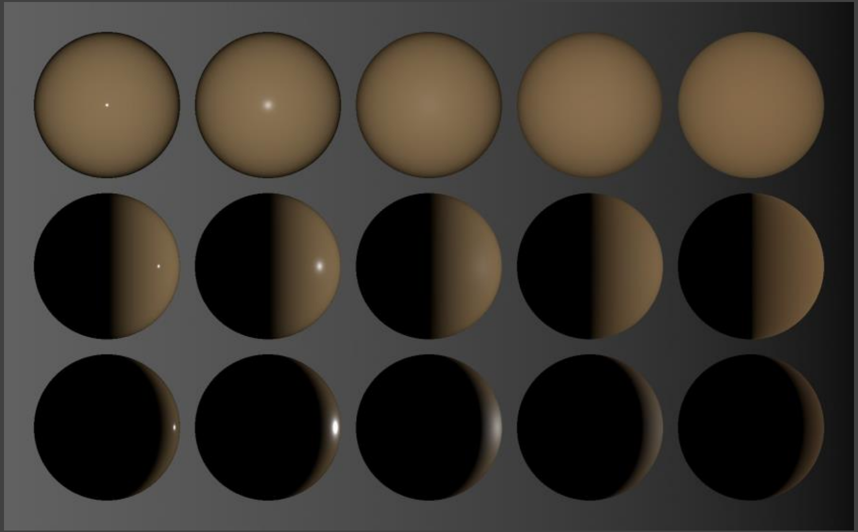
Oren Nayar, $\sigma = 0.5\alpha$

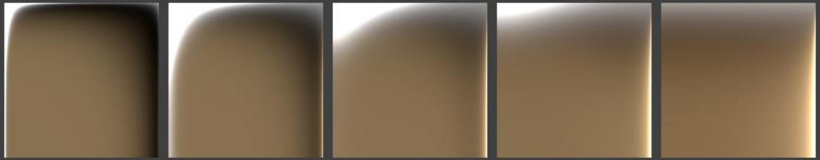


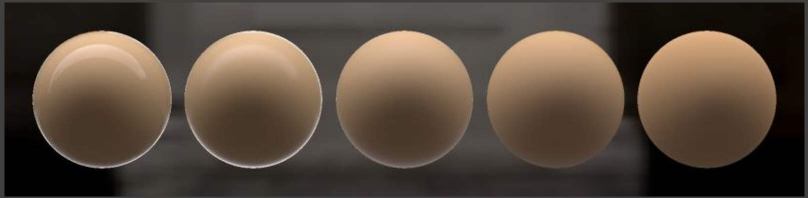


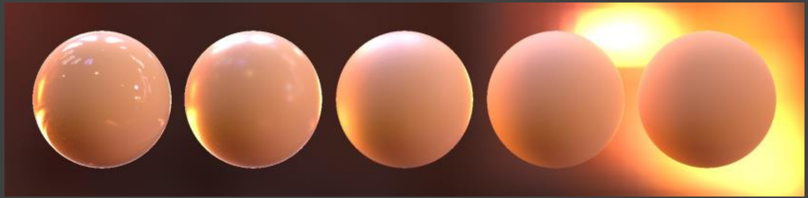


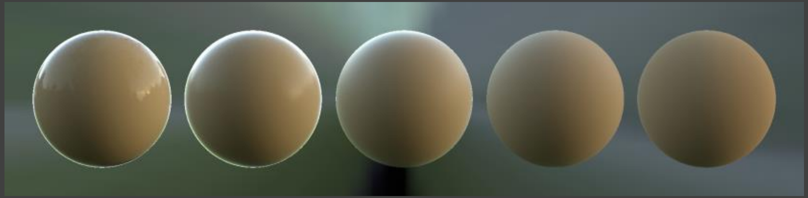
New model

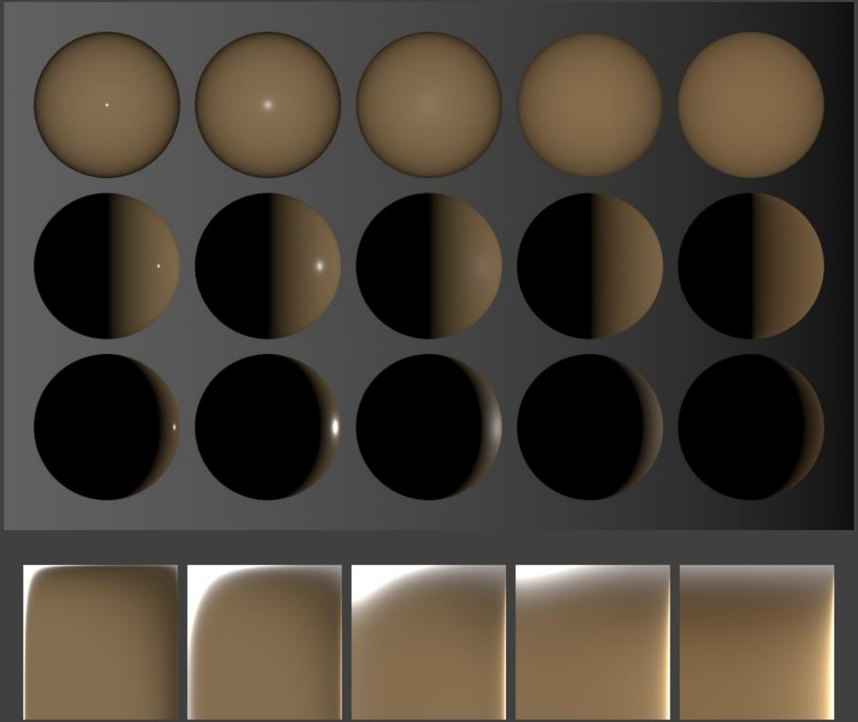






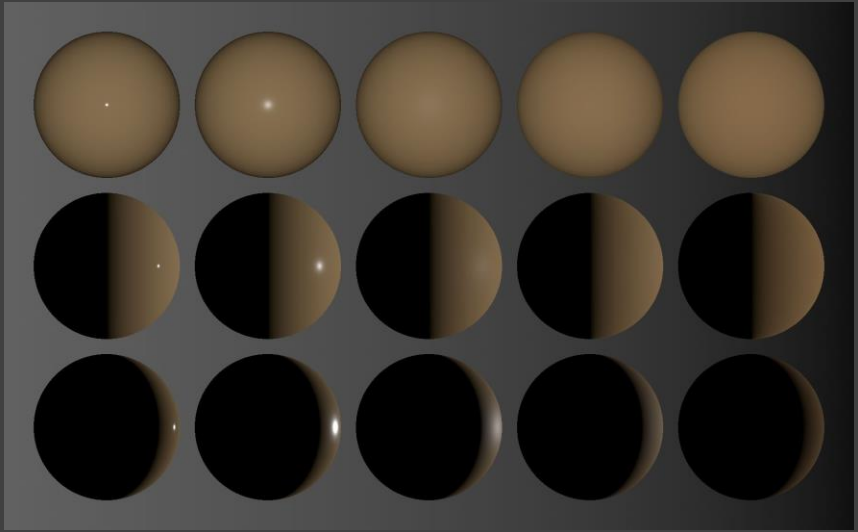


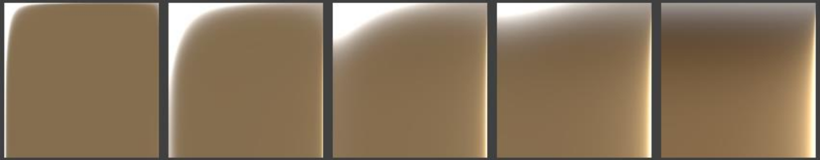




New model (hybrid)

Smooth uses Disney's $f_{d90} = 0.5$, so same as Disney when $\alpha = 0$





New model (cheaper)

Smooth uses Lambert

Lambert

Disney

New model

Lambert

Disney

New model

Lambert

Disney

New model

Special Thanks

- Mark Cerny
  - GDC mentor
- Chad Barb, Xin Liu, Steve Marton
  - Fellow Respawn engineers with awesome ideas/feedback in this research (as always)

Selected References

- Eric Heitz 2014 – Understanding the Masking-Shadowing Function in Microfacet-based BRDFs
  http://jcgt.org/published/0003/02/03/paper.pdf
- Bruce Walter et al 2007 – Microfacet Models for Refraction through Rough Surfaces
  http://www.cs.cornell.edu/~srm/publications/EGSR07-btdf.pdf
- Brent Burley 2012 – Physically Based Shading at Disney
  http://blog.selfshadow.com/publications/s2012-shading-course/burley/s2012_pbs_disney_brdf_notes_v3.pdf
- Brian Karis 2013 - Real Shading in Unreal Engine 4
  http://blog.selfshadow.com/publications/s2013-shading-course/karis/s2013_pbs_epic_slides.pdf
- Peter Shirley et al 1997 - A practitioners' Assessment of Light Reflection Models
  http://www.cs.utah.edu/~shirley/papers/pg97.pdf
- Michael Ashikhman et al 2000 – A Microfacet-based BRDF Generator
  http://www.cs.utah.edu/~shirley/papers/facets.pdf
- Oren Nayar 1994 – Generalization of Lambert's Reflectance Model
  http://www1.cs.columbia.edu/CAVE/publications/pdfs/Oren_SIGGRAPH94.pdf

Questions?

earl@respawn.com

Respawn is hiring!
jobs@respawn.com

Appendix

- The following is a bunch of derivations for Smith shadowing/masking from the ray tracing formulation, and how you use that to actually do the path tracing. This is how I got the results included in the preceding presentation.
- This is quite math heavy. As such, it fits much better in an appendix than in the talk. It is hard to read derivations to an audience, and it is even harder to listen to them! It's better to be able to go at your own pace, and to be able to flip back and forth as needed.
- Final caveat: I didn't polish these appendix slides much (e.g., there is a complete lack of figures). Still, the information and derivation should be helpful to those who like to understand where things come from, and/or who want to understand the Smith shadowing/masking derivation.

Solution – Path Tracing

- Shoot photons into the microsurface for a light direction
- See which view direction those photons come out
- This lets us model diffuse and specular interactions
- But first, we have to be able to ray trace the microsurface
- The microsurface is implicitly defined by the normal distrubution function $D(m)$ and the shadowing/masking function $G(L,V,N)$
- $G(L,V,N)$ derived from $D(m)$ is basically ray tracing
- So, we need to understand how Smith $G(L,V,N)$ works

Starting to derive Smith

- This basically follows Appendix A in Walter's 2007 GGX paper
  - With many missing details filled in, and slightly reordered. Any differences with Walter's appendix are my own attempt to complete the derivation.
  - It may be handy to pull up Walter's appendix as you follow these slides
- Consider ray tracing a 2D slice of the heightfield in the plane of the ray and macrosurface normal
  - Y axis is height (ξ), X axis is projected distance along ray (τ)
- We need probability of hitting height field, given that we haven't yet
- Let $P_1(\xi)$ be the probability density of height ξ
- Probability height ξ is above the heightfield $f(\xi) = \int_{-\infty}^{\xi} P_1(x)dx$
  - This is total probability of heightfield being lower than ξ

PDF of hitting heightfield (1/2)

- For ray $\xi_0 + \mu\tau$ to hit in the next $\Delta\tau$ from height ξ with slope q:
  - $\xi_0 + \mu\,\tau > \xi$
  - $\xi_0 + \mu\,\tau + \mu\,\Delta\tau > \xi + q\,\Delta\tau$
- Rearranging, we have:
  - $\xi_0 + \mu\,\tau > \xi > \xi_0 + \mu\,\tau - (q - \mu)\Delta\tau$
    - Clearly requires $q > \mu$
- Probability of hitting is $\int_{\mu}^{\infty}\int_{\xi_0+\mu\,\tau-(q-\mu)\Delta\tau}^{\xi_0+\mu\,\tau} P_1(x)P_2(q)dx\,dq$
  - $P_1(x)$ is probability density of height x
  - $P_2(q)$ is probability density of slope q
  - Product assumes probability $P_1(x)$ and $P_2(q)$ are independent
    - All normals equally likely at each height; heightfield is fractal, not like canyons or spikes

PDF of hitting heightfield (2/2)

- $\int_{\mu}^{\infty} P_2(q) \int_{\xi_0+\mu\,\tau-(q-\mu)\Delta\tau}^{\xi_0+\mu\,\tau} P_1(x) dx\, dq$
  - Can pull $P_2(q)$ out of the inner integral because it doesn't use x
  - Take $\lim_{\Delta\tau\to 0} \Delta\tau = d\tau$
  - Assume $P_1(x)$ is constant over $d\tau$ at $P_1(\xi_0 + \mu\,\tau)$
- $\int_{\mu}^{\infty} P_2(q) P_1(\xi_0 + \mu\,\tau)(q-\mu) d\tau\, dq$
- Probability of ray $\xi_0 + \mu\tau$ hitting in next $d\tau$:

$$d\tau\, P_1(\xi_0 + \mu\,\tau) \int_{\mu}^{\infty} (q-\mu) P_2(q) dq$$

Meet S, the surviving fraction

- Need conditional probability given we start outside the heightfield:
  - $$\frac{d\tau\, P_1(\xi_0+\mu\,\tau)\int_{\mu}^{\infty}(q-\mu)P_2(q)dq}{\int_{-\infty}^{\xi_0+\mu\,\tau} P_1(x)dx} = \frac{d\tau\, P_1(\xi_0+\mu\,\tau)\int_{\mu}^{\infty}(q-\mu)P_2(q)dq}{f(\xi_0+\mu\,\tau)}$$
- Let $S(\xi,\mu,\tau)$ be the surviving fraction and consider how it changes:
  - $$dS = -\left(d\tau\frac{P_1(\xi_0+\mu\,\tau)\int_{\mu}^{\infty} P_2(q)(q-\mu)dq}{f(\xi_0+\mu\,\tau)}\right)S$$
  - I.e., the fraction of surviving rays hitting in the next $d\tau$ are subtracted from S
- $$\frac{dS}{d\tau} = -\left(\frac{P_1(\xi_0+\mu\,\tau)\int_{\mu}^{\infty} P_2(q)(q-\mu)dq}{f(\xi_0+\mu\,\tau)}\right)S$$

Start solving diff. eq. for S

- We have $\frac{dS}{d\tau} = -\left(\frac{P_1(\xi_0+\mu\,\tau)\int_\mu^\infty P_2(q)(q-\mu)dq}{f(\xi_0+\mu\,\tau)}\right)S$
- In general, $\frac{d}{d\tau}e^{-g(\tau)} = -e^{-g(\tau)}g'(\tau)$, so $S = e^{-g(\tau)}$
  - $g'(\tau) = \frac{P_1(\xi_0+\mu\,\tau)\int_\mu^\infty P_2(q)(q-\mu)dq}{f(\xi_0+\mu\,\tau)}$
    - Recall definition $f(\xi_0+\mu\,\tau) = \int_{-\infty}^{\xi_0+\mu\,\tau} P_1(x)dx$
      - $\frac{df}{d\tau} = P_1(\xi_0+\mu\,\tau)\frac{d}{d\tau}(\xi_0+\mu\,\tau) = \mu\,P_1(\xi_0+\mu\,\tau)$
    - Define $\Lambda(\mu) = \frac{1}{\mu}\int_\mu^\infty(q-\mu)P_2(q)dq$
      - Assumes $\mu \neq 0$
  - $g'(\tau) = \frac{\mu P_1(\xi_0+\mu\,\tau)\frac{1}{\mu}\int_\mu^\infty P_2(q)(q-\mu)dq}{f(\xi_0+\mu\,\tau)} = \Lambda(\mu)\frac{f'(\xi_0+\mu\,\tau)}{f(\xi_0+\mu\,\tau)}$

Final solution for $S(\xi_0,\mu,\tau)$

- $g(\tau) = \int_0^\tau g'(t)dt = \int_0^\tau \Lambda(\mu)\frac{f'(\xi_0+\mu\, t)}{f(\xi_0+\mu\, t)}dt = \Lambda(\mu)\ln f(\xi_0+\mu\, t)|_0^\tau$
  - Uses the fact that $\frac{d}{dt}\ln g(t) = \frac{1}{g(t)}g'(t)$
  - Assumes $f(\xi)$ and $P_1(x)$ are continuous
- $g(\tau) = \Lambda(\mu)(\ln f(\xi_0+\mu\tau)\ \ -\ln f(\xi_0)) = \Lambda(\mu)\ln\frac{f(\xi_0+\mu\tau)}{f(\xi_0)}$
- $S(\xi_0,\mu,\tau) = e^{-g(\tau)} = e^{-\Lambda(\mu)\ln\frac{f(\xi_0+\mu\tau)}{f(\xi_0)}} = \left(\frac{f(\xi_0+\mu\tau)}{f(\xi_0)}\right)^{-\Lambda(\mu)}$
- We've derived the heart of Smith shadowing/masking:

$$S(\xi_0,\mu,\tau) = \left(\frac{f(\xi_0)}{f(\xi_0+\mu\tau)}\right)^{\Lambda(\mu)}$$

Solving for $G_1(\mu)$ given $S(\xi_0, \mu, \tau)$

- We can use this to find the probability of a ray escaping
  - $S(\xi_0, \mu) = \lim_{\tau \to \infty} S(\xi_0, \mu, \tau) = \left(\frac{f(\xi_0)}{f(\infty)}\right)^{\Lambda(\mu)} = \left(\frac{f(\xi_0)}{1}\right)^{\Lambda(\mu)} = f(\xi_0)^{\Lambda(\mu)}$
    - Assumes $\mu > 0$ in $f(\xi_0 + \mu\tau) \to f(\infty)$
- Can find probability of seeing the heightfield in slope $\mu = \frac{d\xi}{d\tau}$:
  - $G_1(\mu) = \int_{-\infty}^{\infty} P_1(\xi)\, S(\xi, \mu) d\xi$
    - Integral over all heights of the probability of having height ξ and escaping in direction μ
  - $G_1(\mu) = \int_{-\infty}^{\infty} f'(\xi)\, f(\xi)^{\Lambda(\mu)} d\xi = \frac{1}{1+\Lambda(\mu)} f(\xi)^{\Lambda(\mu)+1}\Big|_{-\infty}^{\infty}$
    - $f(\infty) = 1$ and $f(-\infty) = 0$ regardless of choice of $P_1(\xi)$, which defines $f(\xi)$
  - $G_1(\mu) = \frac{1}{1+\Lambda(\mu)}$

Solving for $G_2(\mu_L, \mu_V)$

- Can also find visibility in two directions from one height
    - $G_2(\mu_L, \mu_V) = \int_{-\infty}^{\infty} P_1(\xi)\, S(\xi, \mu_L\ ) S(\xi, \mu_V\ ) d\xi$
    - $G_2(\mu_L, \mu_V) = \int_{-\infty}^{\infty} P_1(\xi)\, f(\xi)^{\Lambda(\mu_L)} f(\xi)^{\Lambda(\mu_V)} d\xi$
    - $G_2(\mu_L, \mu_V) = \int_{-\infty}^{\infty} P_1(\xi)\, f(\xi)^{\Lambda(\mu_L)+\Lambda(\mu_V)} d\xi$
- This gives the height correlated Smith shadowing/masking function:
    - $G_2(\mu_L, \mu_V) = \frac{1}{1+\Lambda(\mu_L)+\Lambda(\mu_V)}$
    - Note that $\mu = \cot\theta$, where θ is the angle from the macrosurface normal

Starting to derive $\Lambda(\mu)$

- We still need to finish derivation of $\Lambda(\mu) = \frac{1}{\mu}\int_{\mu}^{\infty}(q-\mu)P_2(q)dq$
- This uses the probability of a tangent slope $P_2(q)$, where $\mathrm{q} = \frac{d\xi}{d\tau}$.
- We have the surface area of a microsurface normal, $D(m)$.
- Project the microsurface normal area onto the macrosurface to get the probability density of a normal per unit area of the macrosurface
  - $D(m)\cos\theta_m$, where $\cos\theta_m = m \cdot N$ is the angle from vertical
- Project from spherical coordinates (θ_m, ϕ_m) to plane $(p, q, 1)$
  - $\frac{(\cos\phi_m \sin\theta_m, \sin\phi_m \sin\theta_m, \cos\theta_m)}{\cos\theta_m} = (\cos\phi_m \tan\theta_m\,, \sin\phi_m \tan\theta_m\,, 1)$

Change variables (θ_m, ϕ_m) to (p, q)

- $p = \cos\phi_m \tan\theta_m,\ q = \sin\phi_m \tan\theta_m$
  - Implies $p^2 + q^2 = \tan^2\theta_m$
- Probability density of normal m is $D(m)\cos\theta_m\,dm$
  - $dm = \sin\theta_m\,d\theta_m d\phi_m$
- We need it as probability density of slopes $P_{22}(p,q)dpdq$
  - This is the same as $D(m)\cos\theta_m\,dm$, just with change of variables.
    - Need the Jacobian, based on partial derivatives
  - $\dfrac{\partial p}{\partial\phi_m} = -\sin\phi_m \tan\theta_m,\ \dfrac{\partial p}{\partial\theta_m} = \dfrac{\cos\phi_m}{\cos^2\theta_m}$
  - $\dfrac{\partial q}{\partial\phi_m} = \cos\phi_m \tan\theta_m, \dfrac{\partial q}{\partial\theta_m} = \dfrac{\sin\phi_m}{\cos^2\theta_m}$

Final Jacobian for (θ_m, ϕ_m) to (p, q)

- $\left|\det\begin{pmatrix} \frac{\partial p}{\partial \phi_m} & \frac{\partial p}{\partial \theta_m} \\ \frac{\partial q}{\partial \phi_m} & \frac{\partial q}{\partial \theta_m} \end{pmatrix}\right| = \left|-\sin\phi_m \tan\theta_m \frac{\sin\phi_m}{\cos^2\theta_m} - \cos\phi_m \tan\theta_m \frac{\cos\phi_m}{\cos^2\theta_m}\right|$
  - This is the Jacobian for the change in area of the measure for a change of variables from (θ_m, ϕ_m) to (p, q)
  - Jacobian = $\left|-\sin^2\phi_m \frac{\tan\theta_m}{\cos^2\theta_m} - \cos^2\phi_m \frac{\tan\theta_m}{\cos^2\theta_m}\right| = \frac{\tan\theta_m}{\cos^2\theta_m} = \frac{\sin\theta_m}{\cos^3\theta_m}$
- This means
  - $\frac{\sin\theta}{\cos^3\theta} d\theta d\phi_m = dpdq$
  - $\sin\theta \, d\theta d\phi_m = \cos^3\theta \, dpdq$

Completing (θ_m, ϕ_m) to (p, q)

- Change of variables for m from (θ_m, ϕ_m) to (p, q) has
  - $dm = \sin\theta_m \, d\theta_m d\phi_m = \cos^3\theta_m \, dpdq$
- We want $P_{22}(p,q)dpdq = D(m)\cos\theta_m \, dm$
  - $P_{22}(p,q)dpdq = D(m)\cos\theta_m \cos^3\theta_m \, dpdq = D(m)\cos^4\theta_m \, dpdq$
- Recall that $p^2 + q^2 = \tan^2\theta_m$… so θ_m is a function of (p, q)
  - If $D(m)$ doesn't depend on ϕ_m, $D(m)\cos^4\theta_m \, dpdq$ is a function on (p,q)!
- For GGX, $D(m) = \frac{\alpha^2}{\pi(\cos^4\theta_m(\alpha^2+\tan^2\theta_m)^2)}$
  - $D(m)\cos^4\theta_m = \frac{\alpha^2}{\pi(\alpha^2+\tan^2\theta_m)^2} = \frac{\alpha^2}{\pi(\alpha^2+p^2+q^2)^2}$

Slope PDF from 2D to 1D

- We're getting close! We have probability density of 2D slope (p, q):
    - $P_{22}(p, q)dpdq$
- We need the 1D probability of slope:
    - $P_2(q)dq$
- Since we've assumed $D(m)$ doesn't depend on ϕ_m, we can arbitrarily rotate (p, q) such that q aligns with the ray direction and p is perpendicular to it
- This lets us integrate $P_{22}(p, q)dpdq$ over all p to get $P_2(q)dq$:
    - $P_2(q)dq = \int_{-\infty}^{\infty} P_{22}(p, q)dpdq$

$P_2(q)$: normals or tangents?

- $P_2(q)$ was derived as the probability density that a microfacet *normal* goes q units along the x-axis (τ) for every 1 unit along the y-axis (ξ).
    - Tangents are always perpendicular to normals.
    - In 2D, vector (x, y) is perpendicular to $(-y, x)$ and $(y, -x)$.
- So, this is equivalent to the microfacet *tangent* slope going $-q$ units along the y-axis (ξ) for every 1 unit along the x-axis (τ).
    - This means $P_2(q)$ is the probability density of tangent slope $-q$.
    - But $D(m)$ doesn't depend on ϕ_m, so $P_2(q) = P_2(-q)$.
- This means $P_2(q)$ is the probability of a microfacet tangent slope q.
    - This is how we used it earlier

Use GGX's $P_{22}(p,q)$ to get its $\Lambda(\mu)$

- For GGX, we saw that $P_{22}(p,q) = \frac{\alpha^2}{\pi(\alpha^2+p^2+q^2)^2}$
- $P_2(q) = \int_{-\infty}^{\infty} P_{22}(p,q)dp = \int_{-\infty}^{\infty} \frac{\alpha^2}{\pi(\alpha^2+p^2+q^2)^2} dp = \frac{\alpha^2}{\pi(\alpha^2+q^2)^{1.5}}$
- All this is to find $\Lambda(\mu) = \frac{1}{\mu}\int_{\mu}^{\infty}(q-\mu)P_2(q)dq$
  - $\Lambda(\mu) = \frac{1}{\mu}\int_{\mu}^{\infty}\frac{\alpha^2(q-\mu)}{\pi(\alpha^2+q^2)^{1.5}}dq = -\frac{\alpha^2+q\mu}{2\mu\sqrt{\alpha^2+q^2}}\Big|_{q=\mu}^{\infty} = -\frac{1}{2}+\frac{\sqrt{\alpha^2+\mu^2}}{2\mu}$
  - We have $\mu = \frac{d\xi}{d\tau}$ for a view vector, so $\mu = \cot\theta_V = \frac{\cos\theta_V}{\sin\theta_V}$ for θ_V from vertical.
- $\Lambda(\mu) = -\frac{1}{2}+\frac{\sqrt{\alpha^2\sin^2\theta_V+\cos^2\theta_V}}{2\cos\theta_V} = \frac{\sqrt{\alpha^2+(1-\alpha^2)\cos^2\theta_V}}{2\cos\theta_V} - \frac{1}{2}$

Smith masking is weird (1/2)

- Ray-tracing derivation of the Smith masking function assumed any height/slope can be immediately adjacent to any other height/slope.
  - I.e., the heightfield is continuous nowhere, yet differentiable everywhere.
    - To be differentiable, you have to be continuous, so this is contradictory.
  - This also means you can’t integrate slopes to get heights, yet slopes are the derivatives of the heights.
    - This is the same contradiction.
  - Smith’s result can be derived just from slope-independent visibility, so there may be a better way to do the ray-tracing derivation.
- We can’t construct a heightfield and just path trace it.
  - Any heightfield with a finite number of heights must violate the assumption of all heights being fully independent.

Smith masking is weird (2/2)

- Ray-tracing derivation has odd result for downward rays ($\mu < 0$).
    - Derivations don't require $\mu > 0$, but do require $\mu \neq 0$.
    - Can show that $\Lambda(-\mu) = -\Lambda(\mu) - 1$.
    - With $\xi_0 < \xi_1$ and $\mu > 0$, we have $S(\xi_0, \xi_1, \mu) = \left(\frac{f(\xi_0)}{f(\xi_1)}\right)^{\Lambda(\mu)}$
    - $S(\xi_1, \xi_0, -\mu) = \left(\frac{f(\xi_1)}{f(\xi_0)}\right)^{-\Lambda(\mu)-1} = \frac{f(\xi_0)}{f(\xi_1)} S(\xi_0, \xi_1, \mu) < S(\xi_0, \xi_1, \mu)$
- Visibility is asymmetric; rays traveling the same path between two heights are more likely to hit something going down than going up!
    - Derivation assumes rays can hit any opposing microfacet normal ($V \cdot m < 0$).
    - More normals oppose downward rays than upward rays.
    - Since the probability distribution of normals is everywhere the same, the cumulative area of candidate normals must be greater for downward rays than for upward rays.
    - In other words, surface area is bigger going down than going up.

Path tracing with Smith masking

- BRDFs must be symmetric
- To be symmetric, we need $S(\xi_1, \xi_0, -\mu) = S(\xi_0, \xi_1, \mu)$
    - Derivation instead has $S(\xi_1, \xi_0, -\mu) = \frac{f(\xi_0)}{f(\xi_1)} S(\xi_0, \xi_1, \mu)$
    - Caused by $\Lambda(-\mu) = -\Lambda(\mu) - 1$
    - To “fix”, redefine $\Lambda(\mu)$ so $\Lambda(-\mu) = -\Lambda(\mu)$, without changing positive slopes:

$$\Lambda(\mu) = \frac{1}{\mu} \int_{|\mu|}^{\infty} (q - |\mu|) P_2(q)\, dq$$

        - Conceptually, renormalize downward surface area to match upward surface area.
    - For GGX, $\Lambda(\mu) = \frac{1}{2}\left(\frac{\sqrt{\alpha^2+\mu^2}-|\mu|}{\mu}\right), \forall \mu \neq 0$

Derivation that $\Lambda(-\mu) = -\Lambda(\mu) - 1$

- This derivation uses the fact that $P_2(-q) = P_2(q)$
- $\Lambda(-\mu) = \frac{1}{-\mu}\int_{-\mu}^{\infty}(\mathrm{q}+\mu)P_2(q)\,dq$
- $-\frac{1}{\mu}\int_{-\mu}^{\mu}(\mathrm{q}+\mu)P_2(q)\,dq - \frac{1}{\mu}\int_{\mu}^{\infty}(\mathrm{q}-\mu+2\mu)P_2(q)\,dq$
- $-\frac{1}{\mu}\int_{-\mu}^{\mu} qP_2(q)\,dq - \frac{1}{\mu}\int_{-\mu}^{\mu} \mu P_2(q)\,dq - \Lambda(\mu) - 2\int_{\mu}^{\infty} P_2(q)\,dq$
- $0 - \int_{-\mu}^{\mu} P_2(q)\,dq - \Lambda(\mu) - \int_{\mu}^{\infty} P_2(q)\,dq - \int_{-\infty}^{-\mu} P_2(q)\,dq$
- $-\Lambda(\mu) - \int_{-\infty}^{\infty} P_2(q)\,dq$
- $-\Lambda(\mu) - 1$

Handling $\mu = 0$

- Preceding derivation assumed $\mu \neq 0$. If you instead assume $\mu = 0$, you get
  - $S(\xi_0, \mu, \tau) = e^{-\tau\left(\frac{P_1(\xi_0)}{f(\xi_0)}\int_\mu^\infty (q-\mu)P_2(q)\,dq\right)}$
    - For GGX, $\int_0^\infty qP_2(q)\,dq = \int_0^\infty \frac{\alpha^2 q}{\pi(\alpha^2+q^2)^{1.5}}\,dq = \frac{\alpha}{2}$, so $S(\xi_0, 0, \tau) = e^{-\tau\frac{\alpha}{2}\frac{P_1(\xi_0)}{f(\xi_0)}}$
- This barely resembles the equation for $\mu \neq 0$
  - $S(\xi_0, \mu, \tau) = \left(\frac{f(\xi_0)}{f(\xi_1)}\right)^{\Lambda(\mu)} = \left(\frac{f(\xi_0+\mu\tau)}{f(\xi_0)}\right)^{-\frac{1}{\mu}\int_\mu^\infty (q-\mu)P_2(q)dq}$
- Limit as $\mu \to 0$ of the equation for $\mu \neq 0$ is the equation for $\mu = 0$!
  - Furthermore, $\lim\limits_{\mu\to0^+}\left(\frac{f(\xi_0+\mu\tau)}{f(\xi_0)}\right)^{-\Lambda(\mu)} = \lim\limits_{\mu\to0^-}\left(\frac{f(\xi_0+\mu\tau)}{f(\xi_0)}\right)^{-\Lambda(\mu)}$
  - True even though $\Lambda(\mu)$ is discontinuous at 0, since $\frac{f(\xi_0+\mu\tau)}{f(\xi_0)} \to 1$.

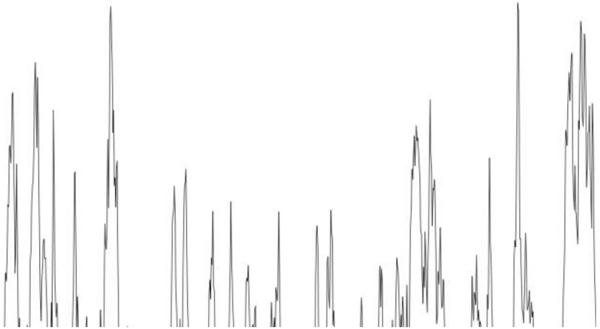
Heightfield heights

- Given $D(m)$, it's possible to figure out the heightfield height limit.
- We have $P_2(q)$, the 1D probability of slope q.
- GGX's cumulative probability $X = \int_{-\infty}^{\mu} \frac{\alpha^2}{2(\alpha^2+q^2)^{1.5}} dq = \frac{\mu}{2\sqrt{\alpha^2+\mu^2}} + \frac{1}{2}$.
- Solve for μ to turn a random variable into a slope: $\mu = \frac{\alpha(2X-1)}{\sqrt{1-(2X-1)^2}}$
- Height is $\sum_i \mu_i d\tau = \sum_i \frac{\alpha(2X_i-1)}{\sqrt{1-(2X_i-1)^2}} d\tau = \alpha \sum_i \frac{(2X_i-1)}{\sqrt{1-(2X_i-1)^2}} d\tau$.
  - All roughnesses can use same heightfield, just scaled by α.
  - Correctly says $\alpha = 0$ is perfectly flat.

GGX heightfield probability func

- I summed random slopes with $\alpha = 1$ to generate heightfields.
  - The random number generator is proven good with a period around 2^{96}.

Number of Slopes	Height Range
2^{16}	± 0.0100
2^{20}	± 0.0030
2^{24}	± 0.0007
2^{27}	± 0.0004
2^{30}	± 0.0002
2^{31}	± 0.0002



- Height histograms were spiky with no correllation between runs.
  - Each run basically picked a random number of random heights to center on.
- Uniform height distribution over $\pm 0.0002\alpha$ seems reasonable.

Starting to path trace GGX

- We now have what we need to path trace GGX:
  - $P_1(\xi) = \begin{cases} \frac{1}{0.0004} & -0.0002 \leq \xi \leq 0.0002 \\ 0 & otherwise \end{cases}$
  - $f(\xi) = \int_{-\infty}^{\xi} P_1(x)dx$
  - $\Lambda(\mu) = \frac{1}{2}\left(\frac{\sqrt{\alpha^2+\mu^2}-|\mu|}{\mu}\right), \quad \mu \neq 0$
  - $S(\xi_0, \mu, \tau) = \begin{cases} \left(\frac{f(\xi_0)}{f(\xi_0+\mu\tau)}\right)^{\Lambda(\mu)} & \mu \neq 0 \\ e^{-\tau\frac{\alpha}{2}\frac{P_1(\xi_0)}{f(\xi_0)}} & \mu = 0 \end{cases}$

Intersection distance (1/2)

- Going from height ξ_0 in direction μ, at what τ do we hit the surface?
- Cumulative probability of hitting the surface is $1 - S$, since S is the cumulative probability of not hitting the surface.
    - We can pick a uniform random variable for $1 - S$ and solve for τ
        - This is equivalent to picking a uniform random variable for S
    - $$S(\xi_0, \mu, \tau) = \begin{cases} \left(\frac{f(\xi_0)}{f(\xi_0+\mu\tau)}\right)^{\Lambda(\mu)} & \mu \neq 0 \\ e^{-\tau \frac{\alpha}{2} \frac{P_1(\xi_0)}{f(\xi_0)}} & \mu = 0 \end{cases}$$
    - $$\tau = \begin{cases} \frac{f^{-1}\left(f(\xi_0)\, S^{-1/\Lambda(\mu)}\right) - \xi_0}{\mu} & \mu \neq 0 \\ -\ln S \frac{2f(\xi_0)}{\alpha P_1(\xi_0)} & \mu = 0 \end{cases}$$
    - $$d = \frac{\tau}{\sqrt{1+\mu^2}}$$

Intersection distance (2/2)

- More convenient to use θ and d.
  - $\mu = \cot\theta \quad \tau = d\sin\theta \quad \mu\tau = d\cos\theta$
  - $\Lambda(\theta) = \frac{\sqrt{\alpha^2+(1-\alpha^2)\cos^2\theta}-|\cos\theta|}{2\cos\theta}, \quad \cos\theta \neq 0$
  - $S(\xi_0,\theta,d) = \begin{cases} \left(\frac{f(\xi_0)}{f(\xi_0+d\cos\theta)}\right)^{\Lambda(\theta)} & \cos\theta \neq 0 \\ e^{-d\frac{\alpha}{2}\frac{P_1(\xi_0)}{f(\xi_0)}} & \cos\theta = 0 \end{cases}$
  - $d = \begin{cases} \left(f^{-1}\left(f(\xi_0)\,S^{-1/\Lambda(\theta)}\right)-\xi_0\right)\tan\theta & \cos\theta \neq 0 \\ -\ln S\,\frac{2f(\xi_0)}{\alpha P_1(\xi_0)} & \cos\theta = 0 \end{cases}$
  - For a 3D vector V, we have $\cos\theta_V = V_z$, making this trivial to calculate

Escaping rays

- If $\cos\theta \neq 0$, then $d = \left(f^{-1}\left(f(\xi_0)\, S^{-1/\Lambda(\theta)}\right) - \xi_0\right) \tan\theta$
- f^{-1} is undefined if $f(\xi_0)\, S^{-1/\Lambda(\theta)} > 1$.
  - Can only happen if $\Lambda(\theta) > 0$, which is when $\cos\theta > 0$ (upward rays).
- Fortunately, algebra shows this is when $S < f(\xi_0)^{\Lambda(\theta)}$.
  - Recall that $f(\xi_0)^{\Lambda(\theta)}$ is the probability of a ray escaping when it starts at ξ_0 and goes in direction θ
  - If $S \leq f(\xi_0)^{\Lambda(\theta)}$, the ray hit the viewer, not the microsurface.
  - This is our one and only path termination condition.

Intersection normal

- Smith derivation assumes $D(m)$ is independent of height.
- So, for a vector traveling in direction T, pick any m according to $D(m)$ such that $T \cdot m < 0$ (ray points at surface, normal points away).
- Need to pick (θ_m, ϕ_m).
  - $D(m) = \frac{\alpha^2}{\pi(1-(1-\alpha^2)\cos^2\theta_m)^2}$
  - Since $D(m)$ doesn't depend on azimuth ϕ_m, just uniformly pick in $[0,2\pi)$.
  - Need to importance sample $D(m)$ over hemisphere
    - $k2\pi\int_0^{\pi/2} D(m)\cos\theta_m \sin\theta_m \, d\theta_m = 1$

Intersection normal

- $Y = 2\pi \int_0^X D(m) \cos\theta_m \sin\theta_m \, d\theta_m$
- $Y = \int_0^X \frac{2\alpha^2 \cos\theta_m \sin\theta_m}{(1-(1-\alpha^2)\cos^2\theta_m)^2} \, d\theta_m$
  - Note that $\frac{d}{d\theta_m} 1-(1-\alpha^2)\cos^2\theta_m = 2(1-\alpha^2)\cos\theta_m \sin\theta_m$
  - We have the form $\int \frac{\left(\frac{\alpha^2}{1-\alpha^2}\right) g'(x)}{g(x)^2} dx$, which has the solution $-\left(\frac{\alpha^2}{1-\alpha^2}\right)\frac{1}{g(x)}$.
- $Y = -\left(\frac{\alpha^2}{1-\alpha^2}\right) \frac{1}{1-(1-\alpha^2)\cos^2\theta_m} \Big|_0^X = -\left(\frac{\alpha^2}{1-\alpha^2}\right)\left(\frac{1}{1-(1-\alpha^2)\cos^2 X} - \frac{1}{\alpha^2}\right)$
- $Y = -\left(\frac{\alpha^2}{1-\alpha^2}\right)\left(\frac{\alpha^2-(1-(1-\alpha^2)\cos^2 X)}{\alpha^2(1-(1-\alpha^2)\cos^2 X)}\right) = -\frac{-1+\cos^2 X}{1-(1-\alpha^2)\cos^2 X} = \frac{1-\cos^2 X}{1-(1-\alpha^2)\cos^2 X}$
  - When $X = \frac{\pi}{2}$ (whole hemisphere), $\cos X = 0$, so $Y = 1$ (i.e., already properly normalized).

Intersection normal

- $Y = \frac{1-\cos^2 X}{1-(1-\alpha^2)\cos^2 X}$
  - Need to solve this for X given Y.
  - $Y - Y(1-\alpha^2)\cos^2 X = 1 - \cos^2 X$
  - $\left(1 - Y(1-\alpha^2)\right)\cos^2 X = 1 - Y$
  - $\cos^2 X = \frac{1-Y}{1-(1-\alpha^2)Y}$
- $X = \cos^{-1}\left(\sqrt{\frac{1-Y}{1-(1-\alpha^2)Y}}\right)$
  - Algebraically equivalent to Walter's result $X = \tan^{-1}\left(\frac{\alpha\sqrt{Y}}{\sqrt{1-Y}}\right)$

Intersection normal

- We can now pick a random microfacet normal:
  - $X_0, X_1 = uniform\ random\ values\ in\ [0,1]$
  - $\phi_m = 2\pi\ X_0$
  - $\theta_m = \cos^{-1}\left(\sqrt{\frac{1-X_1}{1-(1-\alpha^2)X_1}}\right)$
  - $m = (\cos\phi_m \sin\theta_m\,, \sin\phi_m \sin\theta_m\,, \cos\theta_m)$
  - Start over if $T \cdot m \geq 0$.
- Have to retry because there is no closed form solution to importance sample GGX's $D(m)$ directly given the constraint $T \cdot m < 0$.

Intersection normal

- Can be slow to pick a normal as $T_z \rightarrow 1$. Almost none of our guesses satisfy the constraint $T \cdot m < 0$.
- Dot product is $\cos\phi_m \sin\theta_m T_x + \sin\phi_m \sin\theta_m T_y + \cos\theta_m T_z < 0$.
    - $T_x \cos\phi_m + T_y \sin\phi_m < -T_z \cot\theta_m$
- Set $T_x = r\cos\beta$ and $T_y = r\sin\beta$
    - $r = \sqrt{{T_x}^2 + {T_y}^2} = \sqrt{1 - {T_z}^2}; \quad \beta = some\ unknown\ angle$
    - $r\cos\beta\cos\phi_m + r\sin\beta\sin\phi_m < -T_z\cot\theta_m$
    - $\cos(\beta - \phi_m) < -\frac{T_z}{r\tan\theta_m}$; the minimum value for the left is -1

Intersection normal

- $-1 < -\frac{T_z}{r \tan\theta_m}$, so $\tan\theta_m > \frac{T_z}{\sqrt{1-{T_z}^2}}$
- $\cos^2\theta_m < 1 - {T_z}^2$ (trivial since T_z acts like a sine of some angle)
- $\frac{1-X_1}{1-(1-\alpha^2)X_1} < 1 - {T_z}^2$ (the left is our derivation for sampling $\cos^2\theta_m$)
- $1 - X_1 < (1 - {T_z}^2) - (1 - {T_z}^2)(1 - \alpha^2)X_1$
- ${T_z}^2 < \left(1 - (1 - {T_z}^2)(1 - \alpha^2)\right)X_1$
- $X_1 > \frac{{T_z}^2}{1-(1-{T_z}^2)(1-\alpha^2)} = \frac{{T_z}^2}{\alpha^2+{T_z}^2(1-\alpha^2)}$

Intersection normal

- We can now pick a random normal more efficiently:
  - $X_0, X_1 = uniform\ random\ values\ in\ [0,1]$
  - $\phi_m = 2\pi\ X_0$
  - If $T_z > 0$, shrink X_1 to the range $[\frac{{T_z}^2}{\alpha^2+{T_z}^2(1-\alpha^2)}, 1]$
  - $\theta_m = \cos^{-1}\left(\sqrt{\frac{1-X_1}{1-(1-\alpha^2)X_1}}\right)$
  - $m = (\cos\phi_m \sin\theta_m\,, \sin\phi_m \sin\theta_m\,, \cos\theta_m)$
  - If $T \cdot m < 0$, return m
  - If $T_z > 0$, negate ϕ_m. If $T \cdot m < 0$ now, return this m.
  - Start over
- This is more efficient, because at least half the values for X_0, X_1 generate valid normals given the constraint $T \cdot m < 0$

Reflection/Transmission

- Given the microfacet normal m and incoming direction T, we can calculate Fresnel F to decide to reflect or transmit at the facet.
  - $F = F_0 + (1 - F_0)(1 - m \cdot T)^5$, Schlick's famous approximation
  - We use $F_0 = 0.02$, for index of refraction $= 1.33$, common for dialectrics.
- Each ray starts with 1 unit of energy. If the ray's energy is above a threshold, we split it into reflected and transmitted parts with energies scaled by F and $(1 - F)$, respectively. Otherwise, we use Russian Roulette to decide which path gets all the energy.
- Reflected rays continue recursively in the reflection direction.
- Transmitted rays continue in a carefully chosen random direction.
  - If the transmitted ray's energy is above a threshold, we split it into N rays first.

Transmission Direction

- Lambertian scattering would use a cosine weighted hemisphere.
- We tried that. The BRDF was not symmetric.
- The problem is our effective microfacet BRDF was:
  - $F * specular + (1 - F) * diffuse$
  - Both $specular$ and $diffuse$ are valid BRDFs, but F lerps between them based only on the incoming direction $T = -L$.
  - This means that swapping L and V is asymmetric in F; it replaces one vector with an unrelated one.
- In short, Lambertian diffuse doesn't play nicely with a specular BRDF.
  - Fortunately, Shirley et al. solved this in 1997.

Transmission Direction

- Why is Lambertian cosine weighted?
  - Lambertian scattering assumes that light enters the microsurface, bounces around on the inside, and then comes back out.
  - When there is a scattering event inside the surface, it assumes each outgoing direction is equally likely for a ray.
  - You can thus model the interior volume as having uniform beams of energy in every direction. The ones pointing to the surface escape.
  - But the BRDF is defined for a unit area of the surface, not the interior volume. A unit area on the surface cuts diagonally across the uniform beam exiting at an angle, so that the fraction hitting the surface is only $\cos\theta$.

Transmission Direction

- We had a Fresnel reflection on entering the microsurface volume. For symmetry, we need the same Fresnel reflection on exiting too.
- Fortunately, reflection/transmission is symmetric.
- This means we can calculate Fresnel transmission from the view direction into the surface, and it is equivalent to calculating the Fresnel inside the surface for another vector that gets refracted into the view direction.

Transmission Direction

- So, the probability of keeping an exiting direction is $1 - F(\cos\theta_v)$:
  - $1 - (F_0 + (1 - F_0)(1 - m \cdot V)^5)$
  - $(1 - F_0)(1 - (1 - m \cdot V)^5)$
  - We need a normalization constant such that this integrates to 1 over all view directions. That's because rays reflected back into the surface will bounce around and get another chance to escape.
    - Since $1 - F_0$ is a constant, we can just absorb it as part of the normalization constant

Transmission Direction

- $2\pi k \int_0^{\frac{\pi}{2}} (1 - (1 - \cos\theta)^5) \cos\theta \sin\theta \, d\theta = 1$
  - The $\cos\theta$ outside the exponent is the normalization constraint for a BRDF
  - $\sin\theta \, d\theta$ is the measure for integrating over the hemisphere.
  - 2π is from integrating over the hemisphere but not depending on azimuth.
  - k is the normalization constant we want to find.
- This is actually easy to solve. Just multiply out $(1 - \cos\theta)^5$ to get a polynomial in $\cos\theta$. The 1's cancel. We're left with terms like:
  - $a \ \cos^b\theta \sin\theta$
  - Trivial integral of each term is $-\frac{a}{b+1}\cos^{b+1}\theta$

Transmission Direction

- The final result of the integral is:
  - $2\pi k\left(-\frac{5}{3}\cos^3\theta+\frac{5}{2}\cos^4\theta-2\cos^5\theta+\frac{5}{6}\cos^6\theta-\frac{1}{7}\cos^7\theta\right)\Big|_0^{\theta_v}$
  - For $\theta_v=\frac{\pi}{2}$, $\cos\theta_v=0$, and we're left with
    - $2\pi k\left(\frac{5}{3}-\frac{5}{2}+2-\frac{5}{6}+\frac{1}{7}\right)=2\pi k\left(\frac{10}{21}\right)=\frac{20\pi}{21}k=1$
    - $k=\frac{21}{20\pi}=\frac{1.05}{\pi}$
      - Interestingly, this is just 5% larger than the pure Lambertian BRDF.
  - $1+\frac{21}{10}\left(-\frac{5}{3}\cos^3\theta_v+\frac{5}{2}\cos^4\theta_v-2\cos^5\theta_v+\frac{5}{6}\cos^6\theta_v-\frac{1}{7}\cos^7\theta_v\right)$
  - $1-\frac{7}{2}\cos^3\theta_v+\frac{21}{4}\cos^4\theta_v-\frac{21}{5}\cos^5\theta_v+\frac{7}{4}\cos^6\theta_v-\frac{3}{10}\cos^7\theta_v$
    - We want to importance sample this with a [0,1] variable, so can use 1 – this.

Transmission Direction

- $\mathrm{Y} = \frac{7}{2}\cos^3\theta_v - \frac{21}{4}\cos^4\theta_v + \frac{21}{5}\cos^5\theta_v - \frac{7}{4}\cos^6\theta_v + \frac{3}{10}\cos^7\theta_v$
- We can solve this polynomial for $\cos\theta_v$ to use in importance sampling of view directions.
  - No easy closed form solution; follow Shirley’s recommendation to pick a good guess then improve with Newton Raphson.
  - The first guess is $\mathrm{y} = \cos\theta_v = \frac{0.0114813+154.4Y+13002.4Y^2+38295.9Y^3}{1+1483.57Y+33596.4Y^2+16520.2Y^3}$
- We then use this and ϕ_v in $[0,2\pi]$ to get the outgoing transmission direction relative to the microfacet’s normal m.

Transmission Direction

- The result is a combined BRDF:
  - $F * specular + (1 - F) * \frac{1.05\rho_d}{\pi}(1 - (1 - m \cdot V)^5)$
  - The specular part is symmetric because it is nonzero only where $m = H$, which is where $F(m \cdot L) = F(m \cdot V)$
    - $H = normalize(L + V)$, the half-angle vector.
  - The diffuse part is symmetric because $1 - F = (1 - F_0)(\,1 - (1 - m \cdot L)^5)$
    - Swapping L and V just swaps which Fresnel is entering and which is exiting, as expected.
- F is what fraction of rays reflect; $1 - F$ is what fraction transmits.
  - If the ray reflects, only facets aligned to H reflect light, without loss of energy.
  - If the ray transmits, it has to survive another transmission event to escape the surface. Once it escapes, partial absorption has tinted it by ρ_d.

Transmission Observation

- This still doesn't perfectly model subsurface effects
- Most obviously, we ignore how Snell's law changes ray directions
  - Assumptions have ray directions uniformly distributed *inside* the surface
  - Equation assumes rays are uniformly distributed *outside* the surface
  - Snell's law says $\eta_1 \sin\theta_1 = \eta_2 \sin\theta_2$, so $\theta_2 = \sin^{-1}\left(\frac{\eta_1}{\eta_2}\sin\theta_1\right)$
  - θ_2 approximately linear for θ_1 near 0, then θ_2 changes faster as it approaches $\frac{\pi}{2}$ and θ_1 approaches angle of total internal reflection
    - This means outgoing angles are less represented near $\frac{\pi}{2}$ than near 0
    - Conveniently, that's where transmission is weakest, so they're already less represented

Appendix Conclusion

- This now gives all the pieces needed to get the path tracing solution
- I picked about 64 representative zenith angles and 16 α values, and then shot tons of rays for each pair of settings.
- For each ray, I recursively saw what it hit and did Fresnel transmission/reflection based on the chosen normal. Finally, I bucket escaping rays into view directions.
- From this raw data, I tried tons of random equations with terms symmetric in L and V until I saw ones that I liked based on their tradeoff between computation cost and fidelity.